



FACULTAD DE CIENCIAS

GRADO EN BIOLOGÍA

TRABAJO FIN DE GRADO

CURSO ACADÉMICO 2014-2015

TÍTULO:

INGENIERÍA GENÓMICA MEDIANTE SISTEMAS CRISPR-Cas

AUTOR:

YAIZA ARACELI LÓPEZ MANCHEÑO

RESUMEN

Desde su descubrimiento, a finales de la década de los ochenta, hasta la actualidad, las secuencias denominadas CRISPR, se han revelado como instrumentos a través de los cuáles provocar cambios dirigidos en secuencias génicas de interés. El conocimiento de su función y actuación natural en microorganismos permitió a los científicos, sospechar sus posibles aplicaciones como herramienta de edición genómica ya que el sistema era capaz de reconocer elementos genéticos externos, incorporarlos y, tras esto, identificarlos mediante apareamiento de bases y cortarlos. Es este último mecanismo, el de identificación por apareamiento de bases y corte de la secuencia, el que se aprovecha en aplicaciones de ingeniería genómica. La herramienta se compone de dos elementos claramente diferenciados, la fracción nucleotídica constituida, normalmente, por el sgRNA y, la fracción enzimática representada por Cas9. Una vez introducido el sistema en la célula requerida, el sgRNA diseñado de manera específica, guía y señala a la proteína Cas9 la posición de la secuencia a modificar mientras que es esta última la que induce un doble corte en el DNA celular. A partir de este momento, es tarea de los sistemas de reparación celular el introducir la mutación deseada. Se tienen dos vías de reparación, la vía de unión de extremos no homólogos (NHEJ) o la vía de reparación directa por homología (HDR). Mediante la vía NHEJ se promueven mutaciones de inserción o delección que generan, por un cambio en la pauta de lectura, un codón de terminación y, por tanto, una disrupción del gen. Si lo que se desea es modificar la secuencia génica sin truncarla, se debe proporcionar a la célula un molde de reparación portador de la mutación puntual pretendida y se sigue la vía HDR. No obstante, modificando de diferente manera la proteína Cas9, la herramienta es capaz de provocar otro tipo de cambios o de ejercer diversas funciones como la de regulador transcripcional entre otras. Así pues, la importancia del sistema CRISPR-Cas como herramienta de edición genómica, recae en su capacidad para guiar y marcar dianas en el DNA no mediante interacciones DNA-proteína, sino mediante un apareamiento DNA-RNA que promete ser más específico. A través del desarrollo de la técnica de uso y la optimización y depuración de la misma, se prevé que, en un futuro, sea posible la aplicación rutinaria de la tecnología CRISPR-Cas tanto en investigación básica, dónde ya se emplea, como en biotecnología y medicina.

Palabras clave: Modificación genómica; sgRNA; CRISPR-Cas; Cas9.

ABSTRACT

Since its discovery in the late eighties until today, the so called CRISPR sequences have emerged as instrument for creating site-directed changes previously in gene sequences of interest. The knowledge of its function and natural mechanisms inside microorganisms allowed scientists to imagine about possible applications as a genome engineering tool, since the system was able to recognize, incorporate and later on, identify through base pairing and degrade, foreign genetic elements. Is this last mechanism, identifying and degrading by base pairing, the one used in genome engineering applications. The tool consists of two distinct elements, the nucleotidic fraction, constituted normally by sgRNA, and the enzymatic fraction, represented by the protein Cas9. After introducing the system in the required cell, the specifically designed sgRNA guides and targets the Cas9 protein to the sequence position to be modified, which induces a double strand break in the cellular DNA. From this moment, it is the task of cellular repair systems to introduce the desired mutation. There are two main repair pathways involved, the nonhomologous end-joining path (NHEJ) or the homology-directed repair path (HDR). By means of NHEJ pathway, insertion or deletion mutations are promoted and they generate, by a change in the open reading frame, a stop codon and, therefore, a disruption of the gene. If it is desired a modification of the gene sequence without truncating it, a repair template carrying the intended point mutation must be supplied to the cell, in order to follow the HDR pathway. However, by changing the Cas9 in different ways, the tool is able to cause other changes or to exercise various functions including transcriptional regulation, among others. So the importance of the CRISPR-Cas system as a tool for genome edition, lies in its ability to guide and signal targets in the DNA not by DNA-protein interactions, but by a DNA-RNA pairing, which promises to be more specific. Through the development of the usage technique and optimizing or refining it, it is expected that, in the future, the application of the CRISPR-Cas technology both in basic research, where it is already used, as in biotechnology and medicine, may be an everyday routine.

Keywords: Genomic modification; sgRNA; CRISPR-Cas; Cas9.

GLOSARIO DE ABREVIATURAS

Por orden de aparición en el texto:

- ZFNs: nucleasas con dedos de zinc.
- TALENs: nucleasas efectoras tipo activación de la transcripción.
- CRISPR-Cas: agrupaciones de repeticiones cortas palindrómicas regularmente interespaciadas asociadas a nucleasas Cas.
- LTRR: secuencias interespaciadas repetidas en tándem.
- SRSR: repeticiones cortas regularmente interespaciadas.
- pre-crRNA: precursor de RNA transcrito de CRISPR.
- crRNA: RNA transcrito de CRISPR.
- Cas: asociadas a CRISPR.
- PAM: motivo adyacente al protoespaciador.
- tracrRNAs: molécula transactivadora de crRNA.
- sgRNA: RNA guía individual.
- NHEJ: reparación de unión de extremos no homólogos
- HDR: reparación directa por homología.
- WT: silvestre.
- DSB: doble corte en la cadena.
- InDel: inserciones o deleciones.
- REC: lóbulo de reconocimiento.
- NUC: lóbulo con actividad nucleasa.
- HNH: dominio nucleasa.
- RuvC: dominio nucleasa.
- PI: dominio de reconocimiento de la secuencia PAM.
- Cas9n: Cas9 nickasa.
- dCas9: Cas9 defectiva para la actividad nucleasa.
- spCas9: Cas9 de *Streptococcus pyogenes*.
- ChIP: inmunoprecipitación de la cromatina.
- ORF: pauta de lectura abierta.
- ssODN: oligonucleótido de cadena sencilla.
- fCas9: Cas9 con fok I fusionada.
- SNP: polimorfismo de un solo nucleótido.

ÍNDICE

| | |
|--|----|
| 1. Introducción..... | 1 |
| 2. Objetivos..... | 1 |
| 3. Antecedentes..... | 1 |
| 4. Materiales y métodos..... | 6 |
| 4.1 CRISPR-Cas en su estado natural..... | 6 |
| 4.1.1 Qué es CRISPR-Cas..... | 6 |
| 4.1.2 Funcionamiento en su estado natural..... | 7 |
| 4.1.3 Tipos..... | 8 |
| 4.2 Aplicación de CRISPR-Cas como herramienta de edición genómica..... | 10 |
| 4.2.1 Principios generales de ingeniería genómica mediante nucleasas de diseño..... | 11 |
| 4.2.2 Componentes y actuación de CRISPR-Cas9..... | 12 |
| 4.2.3 Aplicaciones como herramienta de edición genómica..... | 17 |
| 4.2.4 Construcción y métodos de introducción..... | 20 |
| 4.2.5 Inconvenientes del sistema CRISPR-Cas9: <i>off target</i> | 23 |
| 4.2.6 Perspectivas de aplicaciones potenciales..... | 26 |
| 4.3 Controversia despertada: patentes e implicaciones éticas..... | 27 |
| 4.4 Simulación práctica: sgRNA..... | 29 |
| 4.4.1 La enfermedad..... | 29 |
| 4.4.2 Diseño del sgRNA..... | 31 |
| 5. Plan de trabajo..... | 33 |
| 6. Conclusiones..... | 35 |
| 7. Bibliografía..... | 36 |

1. INTRODUCCIÓN

En el ámbito de las ciencias biológicas, poco a poco, se ha ido avanzando cada vez a pasos más agigantados en el estudio de la biología molecular y, concretamente, se ha ido profundizando en el campo de la genómica y la genética. En este contexto, el desarrollo de la tecnología del DNA recombinante en la década de los setenta marcó el principio de una nueva era en la biología (Jackson et al. 1972) ya que es a partir de entonces cuando nace la ingeniería genética y, con ella, se impulsa la biotecnología moderna. Gracias a este hito, los científicos podrían manipular el material genético y, así, capacitar el estudio individual tanto a nivel de estructura, organización y función como de regulación, expresión y evolución de un gen. Así pues, hoy en día, somos capaces de mutar de forma dirigida los genes y obtener el cambio deseado en la secuencia que implique, a su vez, el cambio fenotípico específico requerido por el investigador.

Una vez centrados en la ingeniería genética, entendiendo ésta como la aplicación de las tecnologías del DNA recombinante a problemas biológicos, médicos o agrícolas específicos (Pascual Calaforra y Latorre Castillo 2013), aparecen nuevas técnicas significativamente más eficaces y a un coste comparativamente económico que tienen como fin el poder manipular el material genético de cualquier célula. Estas técnicas de edición genómica se basan, en su mayoría, en enzimas nucleasas programables como meganucleasas (Stoddard 2011), nucleasas con dedos de zinc (ZFNs), nucleasas efectoras tipo activación de la transcripción (TALENs) (Urnov et al. 2010) (Scharenberg et al. 2013) y en las agrupaciones de repeticiones cortas palindrómicas regularmente interespaciadas asociadas a nucleasas Cas (CRISPR-Cas) (Hsu et al. 2014; Cox et al. 2015).

2. OBJETIVOS

Así pues, en esta memoria, se describe el estado actual de la ingeniería genómica mediante la herramienta CRISPR-Cas, su importancia, sus bases y aplicaciones y sus perspectivas futuras. A la par, se propone, como vertiente práctica del trabajo, el diseño de uno de los componentes de la herramienta, el RNA guía, con el fin de abordar la reparación de la mutación causante de la anemia falciforme.

3. ANTECEDENTES

Corría el año 1987 cuando se reportó por vez primera la existencia de un conjunto de secuencias repetidas de 29 nucleótidos espaciadas por 32 nucleótidos dentro de un fragmento de DNA codificante para el gen *iap* procedente de una cepa del

microorganismo *Escherichia coli*. En este trabajo, la función propuesta para estos fragmentos fue la de constituir un mecanismo de terminación de la transcripción (Ishino et al. 1987). Siguiendo la estela de estas repeticiones, miembros del mismo grupo de investigación, también confirmaron su presencia en otros microorganismos relacionados como otras cepas de *E. coli*, *Shigella dysenteriae* y *Salmonella entérica* (Nakata et al. 1989). Posteriormente, en el transcurso de los años 90, fueron varios los estudios que reseñaron la existencia de conjuntos de secuencias repetidas e interespaciadas en microorganismos, como diferentes cepas de *Mycobacterium tuberculosis* y las arqueas *Haloferax volcanii* y *Haloferax mediterranei*. En esta última, se encontraron repeticiones en tándem de 30 a 34 pares de bases interespaciadas por secuencias no repetidas de 35 a 39 pares de bases en diferentes regiones de su DNA cromosómico y en su megaplásmido más grande (Mojica et al. 1993; Mojica et al. 1995).

Junto con el desarrollo, desde finales de los años 90 y principios de los 2000, de las nuevas tecnologías de secuenciación y gracias a ellas, fue saliendo a la luz la ubicuidad de las secuencias interespaciadas repetidas en tándem o LTRR (*long tandemly repeated repetitive*) en diversos procariotas bien bacterias bien arqueas. Incluso, se encuentran LTRR en mitocondrias (aunque posteriormente se ha desechado la idea de que estas repeticiones encontradas en mitocondrias jueguen un papel similar al que juegan en procariotas por no poseer el resto de maquinaria asociada a lo que hoy se conoce como CRISPR-Cas) como se recoge en el estudio de Mojica et al. 2000. En él, se constituye una nueva familia de repeticiones de DNA referida como repeticiones cortas regularmente interespaciadas o *Short Regularly Spaced Repeats* (SRSR) y se identifica su presencia en un rango amplio de procariotas (más del 40% de las bacterias secuenciadas y alrededor de un 90% de las arqueas secuenciadas hasta esa fecha). Además de esto, en 2002, se acuñó el término secuencia líder para la secuencia conservada parcialmente localizada en el extremo en el que de forma contigua se sitúan todas las repeticiones cortas interespaciadas (Jansen et al. 2002).

En ese mismo año, tras un consenso entre Jansen y Mojica, se decidió adoptar el acrónimo CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) para estas secuencias repetidas interespaciadas dado el enorme número de acrónimos con los que habían sido nombradas este tipo de secuencias. El acrónimo, actualmente ha quedado fijado y es ampliamente utilizado, aún a sabiendas de que ni todas las secuencias son simétricas en su totalidad ni mucho menos las repeticiones son palíndromos verdaderos en todos los casos sino que, más bien, esto constituye la excepción (Mojica y Garret 2013).

No obstante, aunque se habían hecho progresos en cuanto al descubrimiento de la presencia de CRISPR en multitud de microorganismos, todavía no se sabía de su función ni de la totalidad de sus componentes. A este respecto y también en el año 2002, se constató que de un solo locus de CRISPR se transcribían múltiples RNAs y todos en la misma dirección, por lo que se llegó a la conclusión de que primeramente, se transcribía todo el locus en forma de RNA precursor posteriormente denominado pre-crRNA y, a continuación, este pre-crRNA daba lugar a fragmentos más pequeños de RNA actualmente conocidos como crRNA (Tang et al. 2002). Prácticamente a la par en el tiempo con este estudio, fueron identificados cuatro genes Cas (*CRISPR-associated*) localizados en regiones contiguas a los locus de CRISPR que no existían en microorganismos carentes de los mismos. Por este motivo, se propuso que su funcionalidad debía ir ligada a la de CRISPR (Jansen et al. 2002). Además, en este mismo estudio, se sugirieron las funciones probables de algunas proteínas Cas teniendo en cuenta ciertas homologías con helicasas y exonucleasas.

Aunque llegados a este punto ya se tenía una idea de los componentes que formaban el sistema CRISPR-Cas, todavía no se conocía ni su origen ni su función, esto cambió cuando en el año 2005 se publicaron de manera prácticamente simultánea sendos estudios liderados cada uno de ellos por Mojica y por Pourcel en los que se detallaba la propuesta del origen de las secuencias espaciadoras y, con ello, se planteaba la posible función de CRISPR-Cas. Se descubrió, mediante análisis bioinformáticos, que las secuencias espaciadoras eran homólogas a secuencias pertenecientes a DNA vírico, de cromosomas y de plásmidos que no eran transmisibles. Las secuencias que mejor coincidían con las secuencias espaciadoras se encontraban en elementos genéticos virtualmente capaces de invadir al microorganismo y, por tanto, el origen de estas secuencias estaba en secuencias procedentes de estos elementos genéticos patógenos para el microorganismo que en la actualidad reciben el nombre de protoespaciadores (Mojica et al. 2005; Pourcel et al. 2005). Asimismo se constató que los entes portadores de protoespaciadores ya adquiridos como espaciadores eran incapaces de infectarlos (Mojica et al. 2005), hecho mediante el cual, se determinó la hipótesis de la función de CRISPR como sistema de defensa adaptativo contra elementos de DNA invasor utilizado por microorganismos (Mojica et al. 2005; Pourcel et al. 2005) y que, su mecanismo de actuación era similar al del RNA de interferencia existente en eucariotas (Mojica et al. 2005). A la misma conclusión llegaron Bolotin y sus colegas en un estudio publicado en el mismo año en el que, además, se reseñaba la presencia de una secuencia conservada adyacente al protoespaciador que actuaba como líder (Bolotin et al. 2005) a la que en la actualidad, tras observar su ubicuidad en diferentes sistemas

CRISPR, se denomina PAM (*Protospacer Adjacent Motif*) y que, por lo que se sabe hoy en día, tiene un papel crucial en el funcionamiento de varios sistemas CRISPR-Cas (Mojica y Garret 2013). En referencia a la función desempeñada por el sistema CRISPR-Cas, no fue hasta el año 2007 que se demostró la primera evidencia experimental de su actuación como sistema de inmunidad adaptativa de procariotas, concretamente en *Streptococcus thermophilus*, dado que esta bacteria se volvía resistente a fagos tras adquirir como espaciadores los protoespaciadores presentes en dichos fagos (Barrangou et al. 2007).

Poco a poco, tras estos estudios, se fue profundizando en los mecanismos que sigue el sistema de defensa adaptativo arrojando cada vez más luz al tema. Como por ejemplo la revelación de la actuación de los crRNAs codificados en los espaciadores de CRISPR sobre dianas de DNA (Brouns et al. 2008; Marraffini y Sontheimer 2008), el procesamiento de este crRNA en algunos de los tipos de sistemas CRISPR (Hale et al. 2009; Deltcheva et al. 2011), el corte del DNA diana por parte de la proteína Cas9 guiada por los crRNAs y la importancia que adquiere la presencia de PAMs en estos casos (Garneau et al. 2010) o la necesidad de que los crRNAs se encuentren unidos a tracrRNAs formando estructuras de RNA dúplex para que se asocien con la Cas9 y ésta lleven a cabo su función en el sistema CRISPR-Cas tipo II (Deltcheva et al. 2011).

Se puede decir, que es a partir del año 2011 cuando empiezan a explorarse las aplicaciones en ingeniería genómica del sistema CRISPR-Cas. Para entonces, ya se sabían los tres componentes fundamentales del sistema CRISPR tipo II, la Cas9, el crRNA y el tracrRNA y se pensaba en la manera de reconvertirlo en un sistema nucleasa que pudiera editar genomas guiado por un RNA guía. La primera aproximación fue llevada a cabo por Siksnys y sus colegas que demostraron que el sistema CRISPR-Cas9 podía transferirse y expresarse de manera heteróloga entre bacterias distintas (Sapranauskas et al. 2011). Al año siguiente, mediante la caracterización bioquímica de Cas9 se constató que la enzima purificada podía ser guiada por un crRNA para generar un corte en su correspondiente DNA diana y no sólo esto, sino que fusionando las dos estructuras de RNA del sistema (el crRNA y el tracrRNA) en una construcción abreviada como sgRNA (*single guide* RNA) se veía facilitado el corte del DNA diana *in vitro* (Gasiunas et al. 2012; Jinek et al. 2012). Finalmente, diferentes grupos consiguen, en el año 2013, editar el genoma de células de mamífero mediante la generación de mecanismos de reparación denominados de unión de extremos no homólogos (NHEJ) o reparación directa por homología (HDR) a través de la expresión heteróloga de Cas9 e híbridos crRNA-tracrRNA o sgRNA que dirigen a la proteína hacia su diana de corte o, en el caso de haber expresado varios crRNA distintos, sus respectivas dianas de corte

(Cong et al. 2013; Mali, Yang, et al. 2013). A partir de ese año y hasta la actualidad, se han ido estudiando y analizando de manera más profunda cada uno de los componentes de la herramienta (Jinek et al. 2014; Nishimasu et al. 2014) consiguiéndose así una técnica cada vez más depurada de edición genómica mediante el sistema CRISPR-Cas9 (Figura 1). La herramienta ha sido utilizada para modificar, de varias maneras, los genomas de infinidad de especies como células humanas en cultivo (Cho et al. 2013; Cong et al. 2013; Jinek et al. 2013; Mali, Yang, et al. 2013), bacterias (Wenyan Jiang et al. 2013), nematodos (Friedland et al. 2013), gusanos de seda (S. Ma et al. 2014), ascidias (Sasaki et al. 2014), pez cebra (Hwang et al. 2013), anfibios (Nakayama et al. 2013), roedores (Wang et al. 2013) y plantas (Xie y Yang 2013; Jiang et al. 2013). Y las publicaciones en revistas científicas relacionadas con el empleo de CRISPR en ingeniería genómica van incrementándose año a año e incluso mes a mes siendo en el mes de abril de este mismo año (2015) 83 los artículos que aparecen en el buscador de artículos científicos PubMed (PubMed). Como apunte final para remarcar su importancia, en fecha tan reciente como es el 28 mayo de 2015, las investigadoras Emmanuelle Charpentier y Jennifer Doudna fueron galardonadas con el premio Princesa de Asturias por su contribución en el desarrollo de la herramienta CRISPR-Cas9.

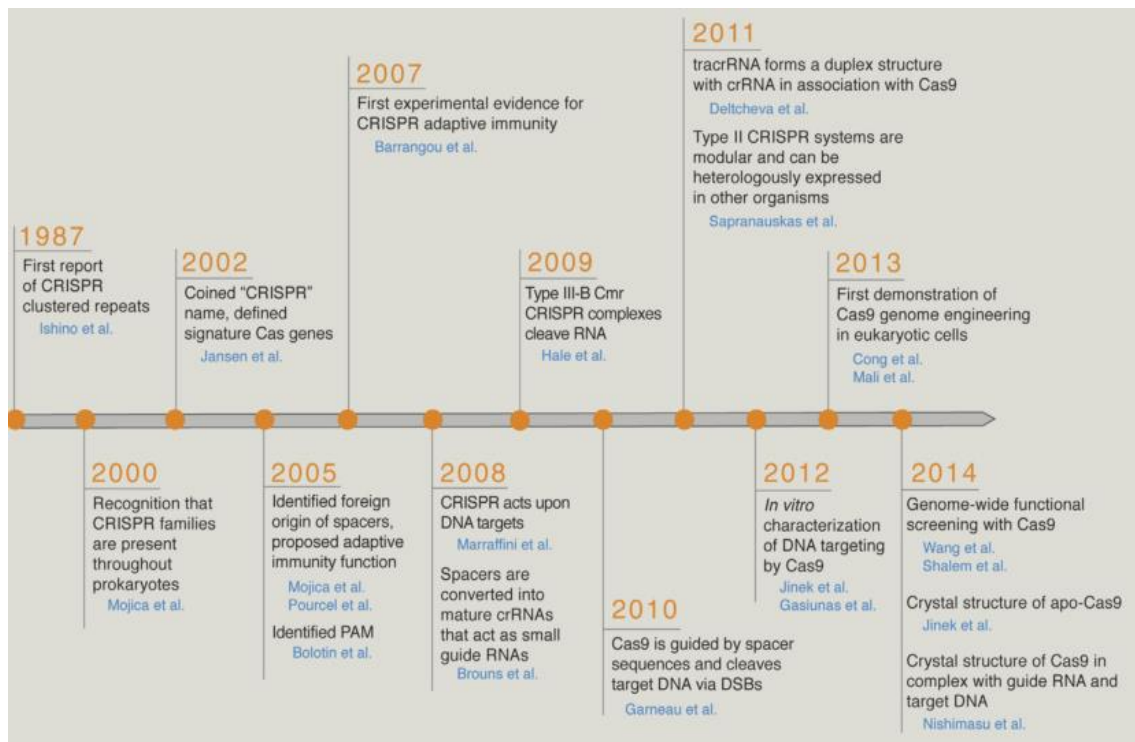


Figura 1 (Hsu et al. 2014). Esquema de la cronología de los avances realizados en referencia al sistema CRISPR-Cas, desde su descubrimiento en 1987 hasta su aplicación como herramienta de edición y manipulación genómica.

4. MATERIALES Y MÉTODOS

Como sabemos, cada carácter fenotípico de cualquier organismo viene determinado en parte por su secuencia génica y, por ende, el mal funcionamiento de cualquiera de sus biomoléculas puede deberse a un cambio o mutación en la codificación original o *wild type* (WT) de sus genes correspondientes. En este sentido, se hace sumamente importante contar con herramientas de edición genómica rápidas, fáciles de aplicar y eficientes, capaces de poner solución de forma irreversible y heredable a la mutación acontecida dentro del gen en cuestión. A este respecto, la herramienta de edición genómica que más relevancia está cobrando de un tiempo a esta parte es CRISPR-Cas cuyo uso, como se explica más adelante, no se ve limitado a la mera generación de mutaciones sino que tiene un rango de actuación mucho más amplio.

4.1 CRISPR-Cas en su estado natural

4.1.1 Qué es CRISPR-Cas

CRISPR-Cas es, en esencia, el mecanismo adaptativo de adquisición de inmunidad y resistencia frente a entes portadores de material genético utilizado por bacterias y arqueas (Ishino et al. 1987; Mojica et al. 2000; Jansen et al. 2002). Se trata de un sistema de dos componentes, CRISPR y Cas estructurados de la siguiente manera:

CRISPR corresponde, como ya se ha referido en párrafos anteriores, al acrónimo de *Clustered, Regularly, Short, Palindromic, Repeat* o agrupaciones de repeticiones cortas palindrómicas regularmente interespaciadas. Este locus contiene los espaciadores adquiridos de virus, plásmidos o cualquier material genético detectado como perjudicial para el microorganismo, las repeticiones que los separan y la secuencia líder en su extremo 5' precediendo a la primera repetición que incluye el promotor de la transcripción.

El locus Cas (*CRISPR-associated*) codifica para proteínas nucleasas cuya secuencia de genes de número variable (3-33) está asociada a CRISPR. Dichas nucleasas, son requeridas para el sistema de defensa multipaso contra los elementos genéticos invasivos (Bhaya et al. 2011).

4.1.2 Funcionamiento en su estado natural

El mecanismo de actuación de CRISPR-Cas como sistema de defensa inmune se puede resumir en tres pasos (Figura 2):

1. Adquisición.
2. Expresión.
3. Interferencia.

En el proceso de adquisición, se reconoce un elemento portador de material genético como agente invasivo, se identifica en su secuencia un fragmento determinado acuñado con el nombre de protoespaciador, se corta y se integra en el locus CRISPR en el extremo 5' tras la secuencia líder y una de las repeticiones cambiando su nombre por el de espaciador. Existen casos en los que en el entorno cercano del protoespaciador, se encuentra el motivo adyacente al protoespaciador o PAM, una secuencia muy conservada cuya función, para con el mecanismo CRISPR-Cas, es lograr que el protoespaciador sea reconocido fácilmente por el sistema de defensa. Tras esto, los espaciadores se expresan en forma de un transcrito de RNA primario denominado pre-crRNA portador de todos los espaciadores del locus que, posteriormente, es cortado por endonucleasas en fragmentos más pequeños (crRNA) conteniendo, cada uno de ellos, un espaciador y una repetición parcial que puede tomar forma de horquilla según el tipo de CRISPR-Cas. De forma ulterior, en la interferencia, cuando el organismo sufre un ataque del agente invasivo, este crRNA acompañado de proteínas Cas entre otras, se une por complementariedad de bases a la secuencia previamente adquirida señalizando a las nucleasas que deben cortar el elemento genético externo. Todos estos acontecimientos, permiten al microorganismo identificar como diana de corte, a elementos extraños portadores de material genético y así, neutralizarlos (Bhaya et al. 2011).

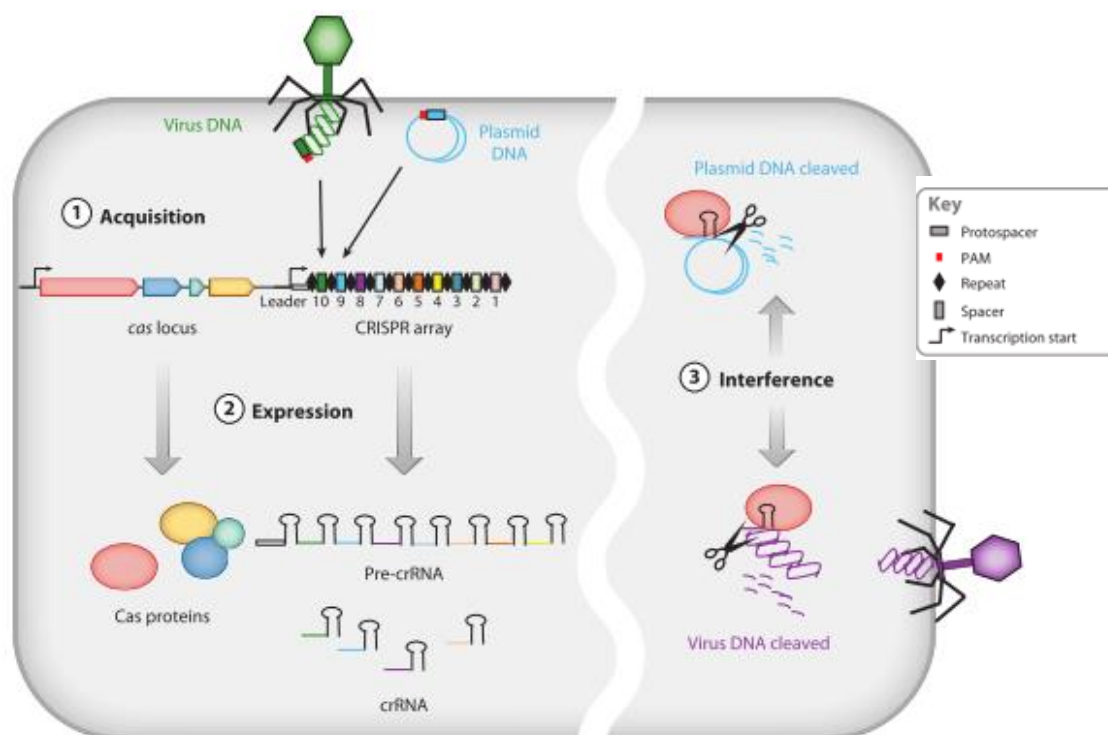


Figura 2 (Bhaya et al. 2011). Ilustración dónde se reflejan las tres fases del mecanismo de defensa de microorganismos llevado a cabo por CRISPR-Cas. 1- Proceso de adquisición dónde se identifica el material genético exógeno y se integra el protoespaciador en forma de espaciador. 2- Expresión de espaciadores primero en forma de precursor y, tras esto, en forma de crRNAs individuales. 3- Interferencia del material genético invasivo mediante la unión del crRNA que señala el corte llevado a cabo por proteínas Cas.

4.1.3 Tipos

Según una clasificación politética basada en la filogenia, conservación de genes, organización de locus y contenido, los sistemas CRISPR-Cas se pueden agrupar en los tipos I, II y III (Makarova, Brouns, et al. 2011) todos ellos conteniendo dos genes hasta ahora pensados como universalmente distribuidos, Cas1 y Cas2. La catalogación de Cas1 y Cas2 como genes universalmente distribuidos es rebatida en una comunicación de tipo personal en la que F. J. M. Mojica expone la no presencia de Cas2 en algunos microorganismos siendo su función asumida por otra biomolécula. En los sistemas CRISPR-Cas poseedores de ambas secuencias estudiados hasta la fecha, Cas1 codifica para una DNAasa y Cas2 para una endorribonucleasa, siendo ambas enzimas metalodependientes y, según parece, de función implicada en el estadio de adquisición (Marraffini y Sontheimer 2009; Yosef et al. 2012). Además de las secuencias de Cas1 y

Cas2, cada uno de los tipos de CRISPR-Cas, posee una secuencia Cas específica (Barrangou y van der Oost 2013) a saber:

- Tipo I → Cas3
- Tipo II → Cas9
- Tipo III → Cas10

No obstante, dentro de los tres tipos, se pueden encontrar subclases basándose, mayoritariamente, en aspectos relacionados con el gen codificante para la Cas1 (Makarova, Brouns, et al. 2011).

Dado el caso de tener que agrupar por similitudes los tres tipos de CRISPR-Cas, se tratarían, por un lado, el tipo I y III por su semejanza tanto bioquímica como estructural, que da lugar a un mecanismo de actuación con una presencia de proteínas Cas relativamente parecida, con las que se desencadena una respuesta en cascada con actividad RNAasa y, por otro lado, el tipo II que, parece haber desarrollado un mecanismo de actuación frente a elementos portadores de material genético que sigue los mismos pasos ya descritos (adquisición, expresión e interferencia) pero de forma sustancialmente distinta a los tipos I y III. En CRISPR-Cas de tipo II, aparece una molécula de RNA transactivador (tracrRNA), transcrita a partir de un locus localizado en trans aguas arriba del locus de CRISPR-Cas, que se une al pre-crRNA y que, de forma posterior, al ser cortado, da lugar a RNA dúplex. Es este RNA dúplex el que se asocia con Cas9 generando así un doble corte o *double strain break* (DSB) en ambas hebras del DNA foráneo y produciendo la destrucción del agente patógeno (Bhaya et al. 2011; Barrangou y Marraffini 2014). En la figura 3, se pueden observar tanto las características compartidas como las disimilitudes entre sistemas.

En referencia a la distribución de los tres tipos de sistemas, se reseña, por el momento, una presencia exclusiva del tipo II en bacterias, la aparición del tipo III más comúnmente en arqueas y la indistinta asignación del sistema tipo I tanto en bacterias como en arqueas (Liu y Fan 2014).

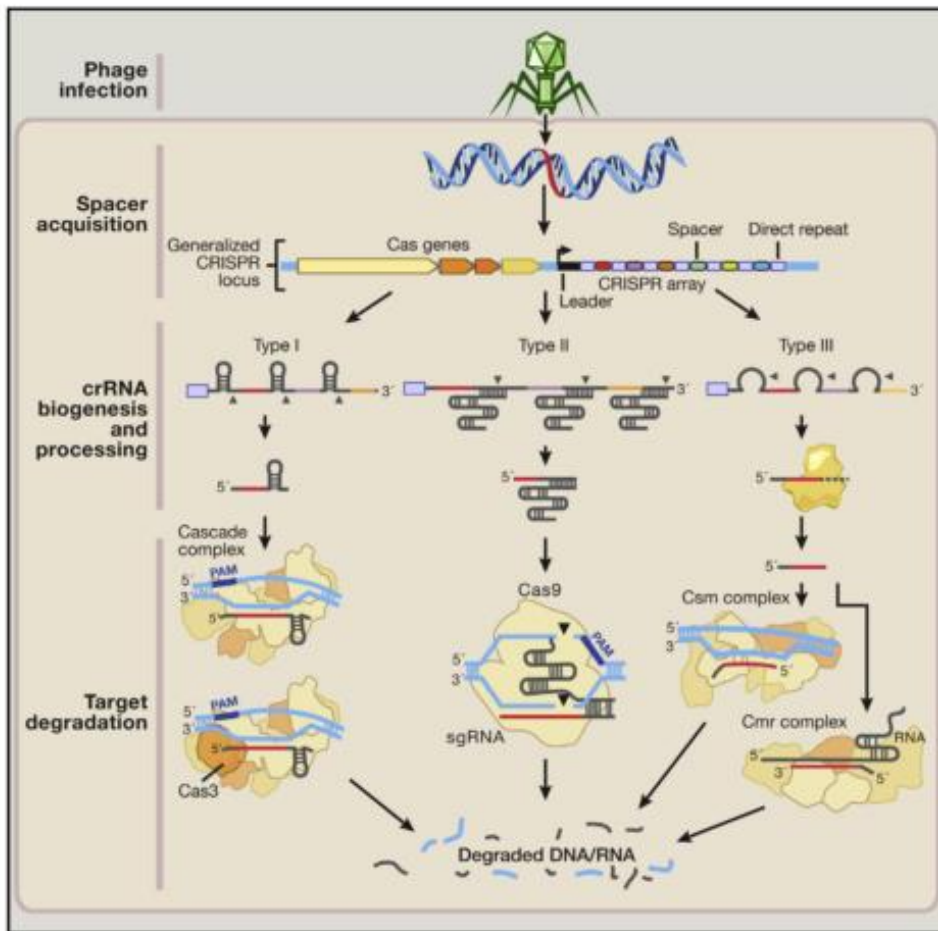


Figura 3 (Hsu et al. 2014). Queda reflejado el distinto comportamiento de los tres tipos descritos de CRISPR-Cas en los procesos de adquisición, expresión e interferencia. Nótese las diferencias más evidentes entre la expresión del crRNA en CRISPR-Cas tipo II y los tipos I y III como son, el empleo del tracrRNA en el tipo II y los complejos formados en los tipos I y III. Otro dato a resaltar es el reconocimiento de secuencias PAM por parte de los tipos I y II y la falta de este reconocimiento del sistema III.

4.2 Aplicación de CRISPR-Cas como herramienta de edición genómica

Una vez revisados los conceptos fundamentales en referencia a los sistemas CRISPR- Cas, en su entorno y funciones naturales, se procede a desarrollar el sistema CRISPR-Cas como herramienta de uso común en el ámbito de la ingeniería genómica en forma de enzimas nucleasas programables, esto es, la utilización del sistema CRISPR-Cas9 .

4.2.1 Principios generales de ingeniería genómica mediante nucleasas de diseño

En primer lugar se hace imprescindible manejar y conocer los principios básicos del proceso de modificación genómica empleando nucleasas de diseño. Dicho proceso, se lleva a cabo mediante dos pasos cruciales ilustrados en la figura (Figura 4).

En el primero de ellos, se debe generar un corte en ambas hebras del DNA bicatenario (DSB) en la diana para la cual se ha diseñado la nucleasa (Urnov et al. 2010). Una vez producido el corte, se pone en marcha la maquinaria celular de reparación de daño en el DNA presente en prácticamente todos los tipos celulares y organismos (Sander y Joung 2014). Esta maquinaria endógena puede seguir dos vías de reparación del DNA, la vía de unión de extremos no homólogos (NHEJ) o la vía de reparación directa por homología (HDR), según se encuentre presente un molde de reparación o no con lo que, según se siga una u otra vía, se conseguirán diferentes tipos de modificación genómica que dependerán del tipo de cambio que el investigador desee promover.

Mediante la vía propensa a errores NHEJ, en la que no se encuentra presente el molde de reparación, los dos fragmentos de DNA bicatenario resultantes del DSB son religados dando lugar, con ello, a modificaciones en forma de inserciones o deleciones (InDel) que, finalmente si se producen en el exón diana, pueden desembocar en *knockouts* por cambio en la pauta de lectura y/o aparición de codones de terminación prematuros (F. Ran et al. 2013).

Si, por el contrario, se introduce un molde de reparación de forma exógena con las alteraciones en la secuencia deseadas, se activa el mecanismo de reparación de daño en DNA vía HDR. Esta vía, implica la recombinación homóloga entre el DNA diana que ha sufrido el ataque nucleofílico y el molde de reparación que puede presentarse en forma de construcción de DNA bicatenario u oligonucleótido de DNA monocatenario, en ambos casos con sendas secuencias de homología con el DNA diana flanqueando la secuencia de inserción (F. Ran et al. 2013). Este mecanismo es capaz de resultar en, potencialmente, cualquier tipo de mutación deseada mediante la sustitución de la secuencia diana por el molde de reparación sintetizado por el investigador con las modificaciones pertinentes (Cong y Zhang 2015).

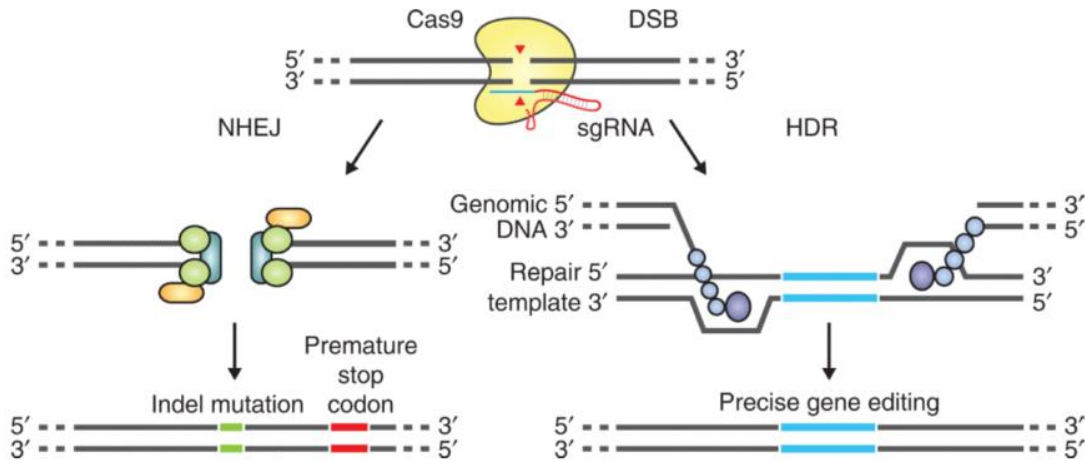


Figura 4 (F. Ran et al. 2013). Los DSB inducidos por Cas9 promueven la edición de genes. Se representan ambas vías de reparación de cortes: la NHEJ tendente a error que puede resultar en mutaciones InDel aleatorias y, por tanto, en cambios en la pauta de lectura generadores de codones de stop prematuros y la HDR que, disponiendo de un molde de reparación, es capaz de dar lugar a una edición precisa y rigurosa en el lugar de corte.

4.2.2 Componentes y actuación de CRISPR-Cas9

Para que el sistema de edición genómica perpetre los cortes en el DNA, se requieren únicamente dos componentes: la proteína Cas9 y el RNA guía cuyas particularidades se detallan a continuación.

La enzima Cas9 se caracteriza por tener una arquitectura bilobular en la que uno de los lóbulos (REC) se encarga del reconocimiento de la diana y el otro (NUC) posee la actividad nucleasa (Nishimasu et al. 2014) (Figura 5 A). Este último lóbulo se compone de dos dominios nucleasa RuvC y HNH, cuyos nombres se basan en la homología a dominios estructurales de nucleasas conocidos, y un dominio de reconocimiento de la secuencia PAM (PI) cuya función se explica más adelante. El primero a reseñar es el dominio similar a RuvC (*RuvC-Like*) que, a su vez viene dividido en tres subdominios desde la región N- terminal donde se localiza el subdominio RuvC I hasta aproximadamente la mitad de la proteína donde se sitúan, a ambos flancos del dominio HNH, los subdominios RuvC II y III (Hsu et al. 2014). El dominio HNH ya mencionado, es semejante al dominio McrA (Cong y Zhang 2015) y, al contrario que RuvC, es un único dominio nucleasa. Ambos dominios, HNH y RuvC, cortan respectivamente la hebra de DNA diana complementaria al RNA guía y la no complementaria al RNA guía

(figura 5 C). Además de esto, el dominio PI localizado en la región C-terminal, parece reconocer la secuencia PAM en la hebra no complementaria al RNA guía situada en el DNA diana (Nishimasu et al. 2014). El lóbulo REC, es el encargado del reconocimiento del otro componente del sistema CRISPR-Cas9 utilizado en ingeniería genómica, el RNA guía. Se divide en tres regiones, hélice puente rica en arginina, REC1 y REC2 (Nishimasu et al. 2014).

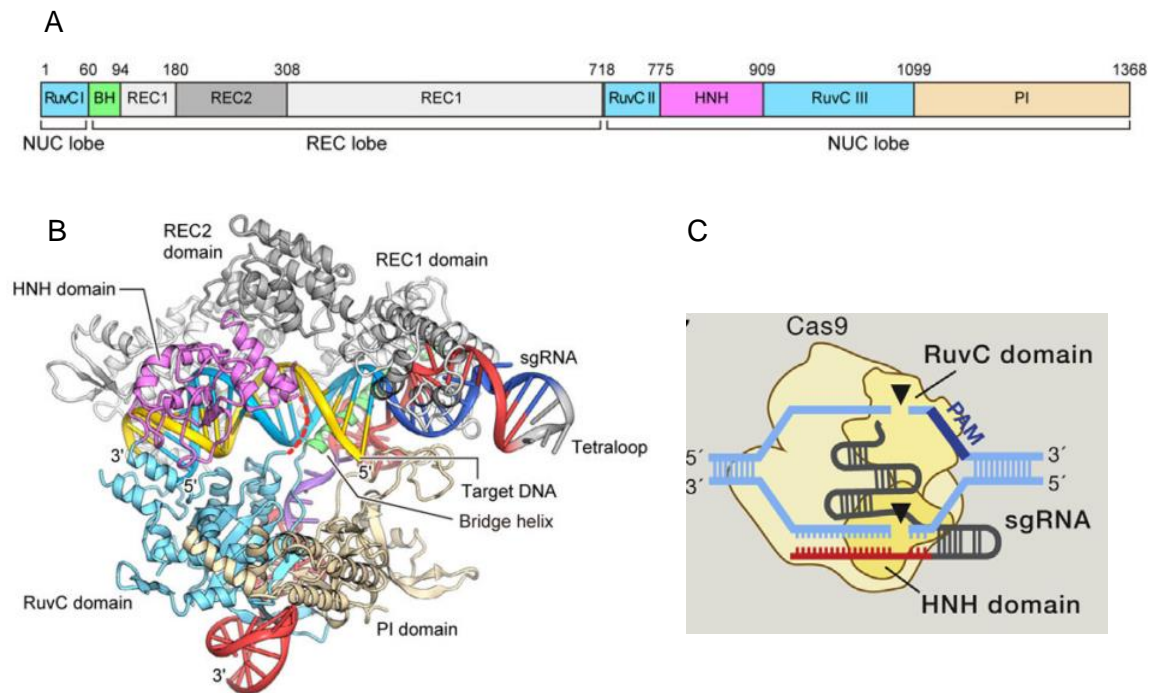


Figura 5 (Nishimasu et al. 2014; Hsu et al. 2014). A- Localización y distribución en el locus correspondiente a la enzima spCas9 de lóbulos REC y NUC con sus respectivos dominios y subdominios. B- Representación gráfica de la estructura del complejo Cas9-sgRNA-DNA. C- Cortes de la hebra de DNA complementaria y no complementaria al sgRNA por parte de los dominios HNH y RuvC respectivamente.

Teniendo en cuenta sendos dominios nucleasa incluidos en Cas9, HNC y RuvC, existen distintas versiones de ésta que pueden ser aplicadas según el tipo de modificación perseguida. Así pues, con Cas9 en su forma silvestre se inducen DSB en la diana mientras que con una versión de Cas9 modificada portadora de uno de sus dominios nucleasa desactivados denominada nickasa (Cas9n) se consiguen mellas o cortes en una sola de las cadenas del DNA diana (Cong y Zhang 2015). Lo que se pretende conseguir mediante la creación de Cas9n es una disminución de cortes en lugares en los que no se desean (*off-target*). Por consiguiente, se introducen dos nickasas, cada una de ellas con uno de sus dominios desactivados siendo estos

diferentes entre sí, que por poseer sendos RNA guía, hacen que la región a identificar en el DNA sea el doble de extensa que con una sola Cas9. Por tanto, consiguen que los requisitos de unión al DNA para producir corte sean mucho más exigentes (Figura 6 B) (Ran et al. 2013) al sumar a la lista de requerimientos de unión una distancia entre ambas enzimas de entre 0 y 20 pares de bases (Davis 2013). Dando un paso más, si ambos dominios catalíticos se desactivan de forma simultánea, esta versión no nucleasa de Cas9 (dCas9) puede ser útil como proteína de unión a DNA guiada por RNA dotando al sistema de aplicación como modulador transcripcional dirigido (Perez-Pinera et al. 2013).

Existen, dentro del conjunto de proteínas Cas9, distintos ortólogos, en cuanto a la extensión de su secuencia aminoacídica, encontrándose tamaños de entre 900 y 1600 aminoácidos. Este hecho es independiente y no determina la subclasificación dentro de CRISPR-Cas tipo II. La subclasificación se divide en IIA, IIB y IIC, y se basa en la presencia o ausencia de distintas Cas en el locus (Makarova, Aravind, et al. 2011). El tipo IIC posee el conjunto de Cas básico, Cas9, Cas1 y Cas2 mientras que el tipo IIA añade Csn2 y el tipo IIB Cas4 (Figura 6 A) (Chylinski et al. 2013). La diferencia en la extensión de Cas9, recae en el lóbulo REC. A pesar de esto, la arquitectura de sus dominios es muy similar y comparten regiones altamente conservadas en sus dominios catalíticos (Figura 6 C). Según la bibliografía consultada, el subtipo de sistema CRISPR-Cas9 más utilizado con fines ligados a la ingeniería genómica es el IIA de *Streptococcus pyogenes* también abreviado como spCas9 (Dicarlo et al. 2013; Hwang et al. 2013; Jinek et al. 2012; Konermann et al. 2014; Larson et al. 2013; Maeder et al. 2013; Paix et al. 2014; Perez-Pinera et al. 2013; F. A. Ran et al. 2013; F. Ran et al. 2013; Schwank et al. 2013; Shao et al. 2014; Wang et al. 2013; Shen et al. 2014; Yang et al. 2014).

La otra mitad del sistema CRISPR-Cas9 de diseño consiste en un componente de RNA dúplex ya mencionado anteriormente formado por crRNA y la molécula de crRNA transactivador tracrRNA, denominado de forma general RNA guía. El crRNA complementa por apareamiento de bases con su diana en el DNA o protoespaciador a lo largo de 20 nucleótidos. Es preciso aclarar que en este apareamiento de bases no tiene por qué darse la complementariedad perfecta. En el extremo 3' de estos 20 nucleótidos se halla una repetición que, a su vez, aparea con el otro constituyente del RNA guía, el tracrRNA formando una estructura muy característica con varias horquillas. Se sabe que la molécula de tracrRNA es requerida para la unión del RNA guía con la Cas9 y la buena orientación de éste para con la interacción correcta entre el crRNA y su diana haciéndose indispensable su presencia para el funcionamiento de la herramienta de edición genómica (Jinek et al. 2012).

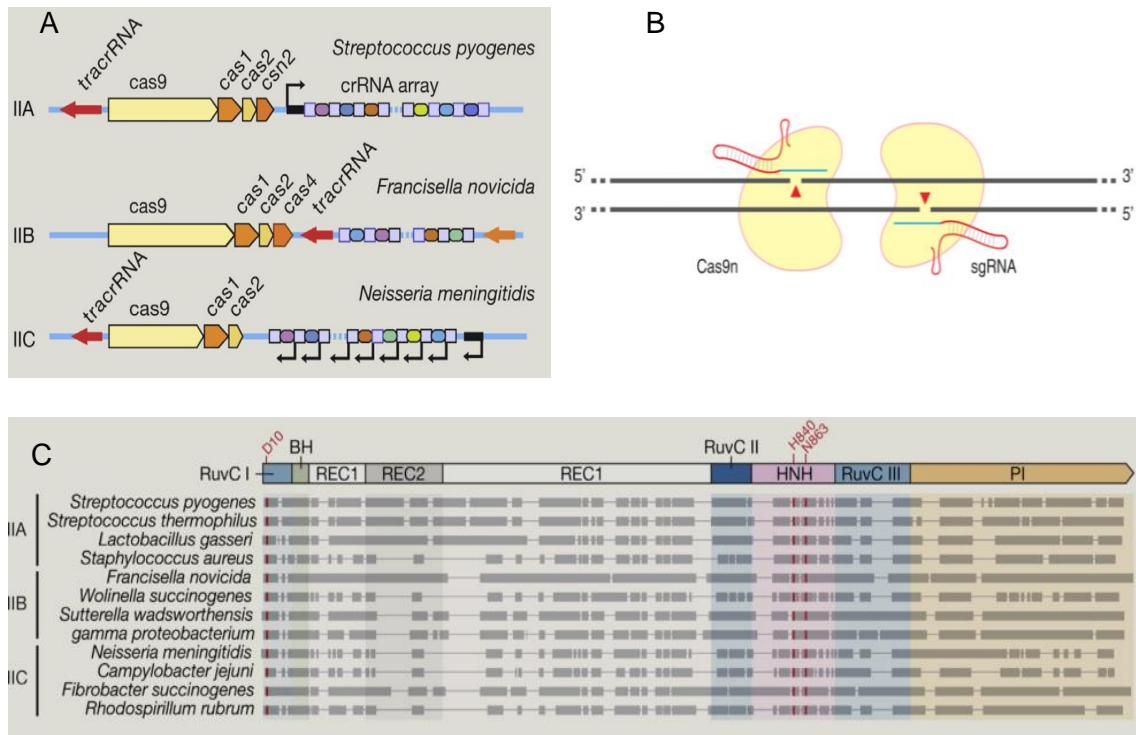


Figura 6 (Hsu et al. 2014; F. A. Ran et al. 2013). A- Organización en los locus de los subtipos de CRISPR tipo II (IIA, IIB y IIC) clasificados según sus secuencias de genes Cas. B- Generación de DSB mediante el uso de dos nickasas con uno de sus dominios inactivados. C- Extensión de la secuencia Cas9 en diferentes microorganismos, similitudes estructurales entre sus dominios y conservación de secuencias dentro de los dominios catalíticos (en rojo).

El crRNA:tracrRNA, puede construirse en forma de dúplex de igual manera en la que podemos encontrarlo en su estado natural o bien puede fusionarse creando una quimera en forma de una sola cadena con un lazo de unión entre sus dos elementos constituyentes cambiando su denominación por la de sgRNA, abreviatura de *single guide RNA*. Este sgRNA sigue teniendo las características clave del RNA guía, la secuencia de 20 nucleótidos en su extremo 5' que hibrida con el protoespaciador presente en el DNA y la estructura bicatenaria en su región 3' que facilita la unión entre el sgRNA y la Cas9 (Figura 7) (Jinek et al. 2012; Doudna y Charpentier 2014). Además de esto, el sgRNA ha conseguido demostrar una eficiencia incluso superior comparado con el crRNA:tracrRNA (Cong y Zhang 2015) la cual podría aumentarse más si cabe al acortar el sgRNA de 20 nucleótidos a 17 nucleótidos según demuestra un estudio (Fu et al. 2014).

Aunque siendo rigurosos, la secuencia PAM como tal no forma parte, estructuralmente hablando, de la herramienta CRISPR-Cas9, es imposible dominar ésta sin conocer la implicación de su presencia en el proceso ya que es determinante. La secuencia PAM, cuyo acrónimo, como ya se ha referido anteriormente, deriva de los términos en inglés *Protospacer Adjacent Motif* acuñados por Mojica et al. en 2009, se compone de una serie de nucleótidos localizados en la región 3' de la hebra de DNA diana que no se une al sgRNA a pocos nucleótidos de la secuencia diana (Figura 7). La presencia de esta secuencia, implica el reconocimiento por parte de Cas9 de la diana de corte pues la endonucleasa con su sgRNA asociado primero reconoce a PAM produciéndose la unión DNA-Cas9, posteriormente se separan las dos hebras y la herramienta se desplaza aguas arriba de la hebra en busca de su diana (Sternberg et al. 2014). Además de esto, junto con la unión del sgRNA por complementariedad de Watson y Crick con el DNA, la secuencia PAM parece determinar la transición entre la actividad de Cas9 de reconocimiento y unión a la de corte (Sternberg et al. 2014; Jinek et al. 2014; Nishimasu et al. 2014). Es importante recalcar que el reconocimiento de la secuencia PAM se lleva a cabo, como se ha mencionado anteriormente, a través del dominio PI de Cas9 y no mediante el sgRNA.

Así, la secuencia de acontecimientos que dan lugar al corte en el DNA diana sería la siguiente, reconocimiento de la secuencia PAM en la hebra de DNA no complementaria al sgRNA, unión de Cas9, desplazamiento aguas arriba a fin de que el sgRNA localice la secuencia diana, localización de la secuencia diana mediante la unión por complementariedad de bases y acomodación del heterodúplex sgRNA:DNA diana en el surco de carga positiva alojado entre los lóbulos REC y NUC, activación de los dominios nucleasa HNC y RuvC y corte del DNA posicionado con los grupos fosfato escindibles correctamente a 3 o 4 nucleótidos de distancia de PAM (si PAM se encuentra lejos de la diana no se produce el corte).

La secuencia nucleotídica de PAM, puede variar dependiendo del microorganismo del que proceda la Cas9 utilizada o de posibles modificaciones que se realicen en su dominio PI. Algunas de esas PAM para diferentes ortólogos de Cas9 son 5'-NGG de SpCas9, 5'-NAG para una SpCas9 modificada, 5'-NNAGAAW para CRISPR1 de *Streptococcus thermophilus*, 5'-NGGNG para CRISPR3 del mismo microorganismo o 5'-NNNNGATT para la Cas9 de *Neisseria meningitidis*. La utilización de distintas Cas9 con diferentes PAM permite al experimentador, por ejemplo, modificar varios sitios del locus de una célula de manera independiente y simultánea (Hsu et al. 2014).

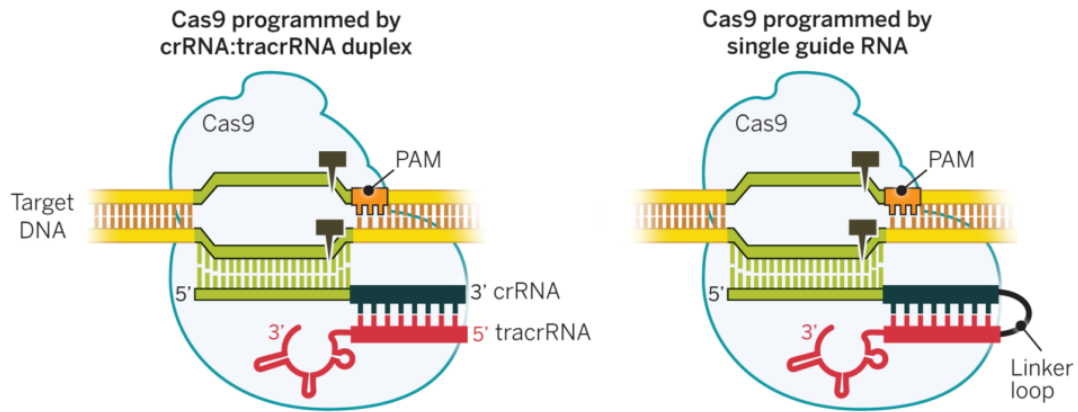


Figura 7 (Doudna y Charpentier 2014). A la izquierda el RNA guía de estructura dúplex formado por el crRNA y el tracrRNA. A la derecha, el sgRNA formado únicamente por un RNA resultante de la fusión mediante un lazo de unión entre el crRNA y el tracrRNA. Nótese también la presencia y posición de la secuencia PAM.

4.2.3 Aplicaciones como herramienta de edición genómica

CRISPR-Cas9 en su uso como herramienta de edición genómica, puede ser programada para desempeñar distintas funciones con diferentes efectos a saber, principalmente, interrupciones mediante inserciones y deleciones (InDel), modificaciones o ediciones endógenas y activación o represión de la expresión génica, todas ellas desarrolladas brevemente a continuación. No obstante, al margen de su utilización como herramienta de edición genómica, también se puede emplear en técnicas de purificación de regiones de DNA genómico en combinación con inmunoprecipitación de la cromatina (ChIP), marcaje de regiones de DNA genómico para imagen y escrutinios genómicos con genotecas agrupadas.

Como ya se ha citado anteriormente, para que se produzca una interrupción de un gen en forma de deleción o inserción mediante el empleo de Cas9 y su correspondiente sgRNA, se debe generar un DSB cuya reparación se lleve a cabo mediante la vía NEJ carente de molde de reparación. A través de ésta y tras el DSB, un bajo número de nucleótidos pueden ser insertados en el DNA o deleccionados de éste de manera aleatoria en el sitio de corte o diana. Estas pequeñas inserciones o deleciones alteran la pauta abierta de lectura (ORF) aguas abajo de la secuencia diana contenida en el gen lo cual producirá un cambio en su secuencia aminoacídica que, tarde o temprano, resultará en la aparición de un codón de *stop* o terminación prematuro truncando así el transcrito de ese gen y generando una pérdida de función (Figura 8 A). Teniendo en cuenta el resultado que se busca, una interrupción, el sgRNA ha de diseñarse para dianas

posicionadas tan cerca como sea posible a la región N-terminal, es decir, lo más al principio posible del inicio de la región codificante del gen que se pretende interrumpir ya que, de esta manera, se asegura la pérdida total de función del mismo. Debido a que los InDel son aleatorios se hace necesario que el investigador, tras llevar a cabo la disrupción, determine de forma experimental el tipo y extensión de la misma (Addgene) siendo, en todo caso, modificaciones permanentes e irreversibles, afirmación que atiende a la lógica teniendo en cuenta que es una secuencia de DNA genómico lo que se muta.

De forma contraria al mecanismo de disrupción génica mediante Cas9 y como ya se ha referido en párrafos anteriores brevemente, para obtener la reparación o cambio en determinado nucleótido previamente definido del gen de interés, tras la generación de un DSB no debe seguirse la vía de reparación NHEJ sino la de Reparación por Homología Directa (HDR) (Figura 8 B) ya que la primera induce reparaciones imperfectas e indefinidas. La presencia de un molde de reparación, normalmente diseñado en forma de oligonucleótido de cadena sencilla (ssODN), con la mutación puntual que se desee introducir es indispensable para generar la vía HDR. Este molde, se inserta en la célula junto con la Cas9 y el sgRNA y debe cumplir una serie de características muy exigentes. A fin de que la mutación puntual se genere en la posición correcta del gen, el molde debe tener un alto grado de homología con la secuencia de DNA endógeno en sus regiones inmediatamente aguas arriba y aguas abajo del DSB es decir, el lugar de homología debe encontrarse a menos de 100 pares de bases o, idealmente, 10 pares de bases de distancia del sitio de corte. Otros requerimientos a cumplir por el molde dependen de la extensión de nucleótidos que se pretenda mutar. Así, se utilizará un molde en forma de ssODN si el cambio a introducir no es de un tamaño superior a 50 pares de bases ya que es el diseño en el que resulta más eficiente, en este caso los brazos de homología que flaquean a la mutación tendrán una longitud aproximada de entre 50 y 80 bases cada uno. En cambio, para cambios de tamaño superior a 100 pares de bases se usa un plásmido en el que los brazos flanqueantes tienen una extensión de alrededor de 800 bases, siendo este tipo de modificaciones menos eficientes. También difiere, según la longitud del cambio previsto, la concentración de ssODN o plásmido que se inserta en la célula. Asimismo, es de extrema importancia la supervisión del molde en busca de secuencias PAM específicas de la Cas9 que se emplea pues su existencia resultaría de forma probable en la identificación del molde como diana por parte del sgRNA y, por ende, de su corte enzimático. Para evitar que esto ocurra, si se encuentran PAMs en el molde de reparación, éstas se mutan teniendo en cuenta que dicho cambio de lugar a una

mutación silenciosa en el caso de que la PAM se halle en la región codificante. Es decir si, por ejemplo, se utiliza la spCas9 se buscarán secuencias que correspondan a NGG que, en el caso de que se encuentren, se modificarán a NGT, NGA o NGC. De esta forma, se pueden introducir mutaciones que van desde 1 nucleótido hasta 1 kilobase (según referido por el laboratorio de Zhang). Al igual que en el caso de las interrupciones génicas, el científico debe confirmar mediante métodos experimentales la modificación deseada en el DNA genómico. Cabe mencionar, también, que las modificaciones en forma de interrupción génica tienen una eficiencia superior a las modificaciones que tienen como objetivo el reparar o modificar de forma específica la secuencia nucleotídica (Addgene).

Por último, si el uso que se le quiere dar a la herramienta CRISPR-Cas9 es el de modulador de la transcripción génica ya sea activándola o reprimiéndola, se precisa una dCas9 con ambos dominios nucleasa desactivados cuya especificidad de unión a la diana no se ve alterada (Figura 8 C y 8 D). Esta dCas9, normalmente, se funcionaliza con diferentes dominios efectores fusionados que, en el caso de perseguir un resultado de activación transcripcional, serán dominios activadores de la transcripción y, en el supuesto de buscar una represión, existirán dos opciones, el empleo de la dCas9 en solitario o, para conseguir un efecto *knockdown* superior, la fusión de la enzima a dominios silenciadores de la transcripción. Para poder asegurar con mayor probabilidad una activación o represión, el diseño del sgRNA debe hacerse teniendo en cuenta que su diana de unión se encuentre lo más cerca posible del promotor del gen, facilitando así la actuación del activador o silenciador transcripcional fusionado a dCas9 (Addgene). Así, una vez la enzima se une a su diana, puede tanto bloquear el inicio de la transcripción (si su diana de unión es el promotor de la secuencia génica o sus regiones adyacentes implicadas en la unión de la maquinaria de transcripción) o elongación (Larson et al. 2013) como reclutar la maquinaria celular de transcripción para aumentar la expresión de determinado gen según se busque un resultado de activación o represión transcripcional respectivamente. La acción de dCas9 con dominios efectores es modulable tanto en cuanto su expresión dentro de la célula puede ponerse bajo el control de un promotor inducible (Gilbert et al. 2014). Además de esto, parece ser que fusionando no sólo un dominio efector sino tres, se consiguen niveles de transcripción mayores (Chavez et al. 2015). Esta aplicación de CRISPR-Cas como modulador de la transcripción, al contrario que la de reparación o la de interrupción, no produce cambios definitivos ni permanentes dado que en ningún momento afecta o modifica de forma directa la secuencia del gen al que regula.

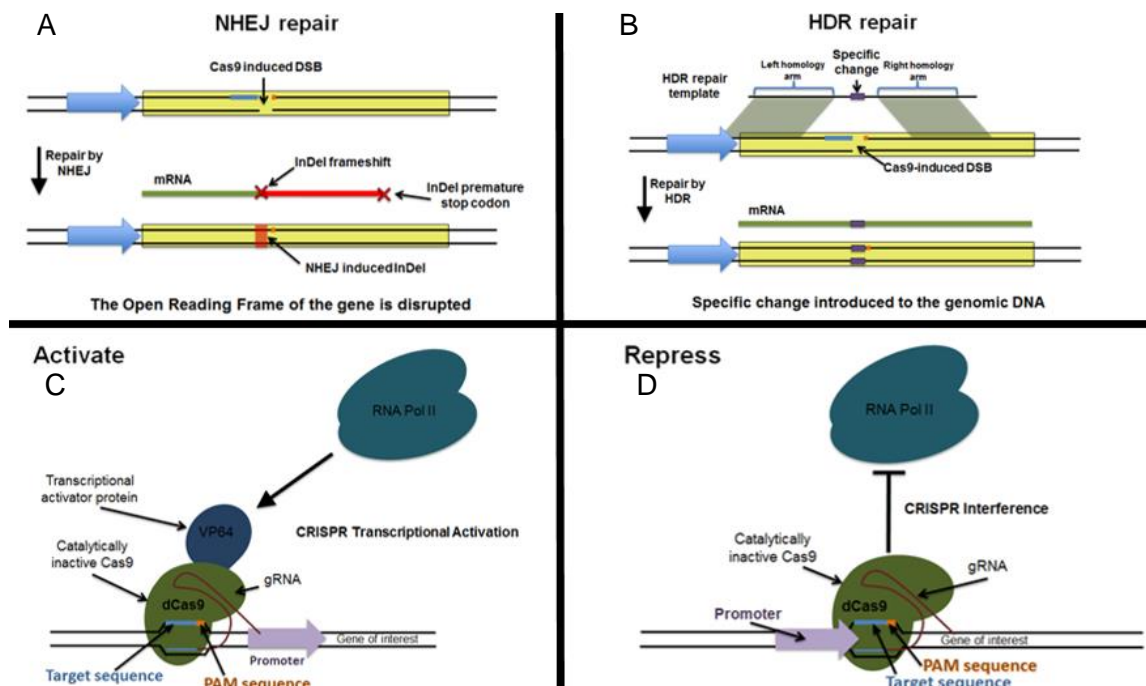


Figura 8 (Addgene). Representación de las distintas funciones principales que puede desempeñar la herramienta CRISPR-Cas9. A- Introducción de mutaciones tipo InDel. B- Generación de modificaciones tipo *repair*. C- Activación de la expresión génica. D- Represión de la transcripción génica.

4.2.4 Construcción y métodos de introducción

De manera usual, para construir vectores de expresión conteniendo en su secuencia el sgRNA y la Cas9 se recurre a plásmidos de una de las distintas compañías biotecnológicas. Las secuencias oligonucleotídicas de ambos componentes de la herramienta tan solo han de ser insertadas en forma de DNA, en el caso del DNA codificante para el sgRNA con un paso previo de alineamiento, en los vectores y éstos, a su vez, introducidos en las células mediante procedimientos que se detallarán más adelante. Los dos componentes, la Cas9 y el sgRNA en forma de DNA, pueden contenerse en un solo vector de expresión o en dos de ellos por separado si bien es cierto que la tendencia actual es introducirlos en un mismo vector (F. Ran et al. 2013; Yang et al. 2014; Shao et al. 2014). Si la opción elegida es introducir ambos elementos en el mismo vector de expresión, el que se emplea de manera habitual es pX330, desarrollado por el laboratorio de Feng Zhang, el cual tiene como promotor de expresión de sgRNA a U6 (un promotor humano) y de la nucleasa a un híbrido de beta actina procedente de pollo (CBh) (Figura). Además de la capacidad de portar sendos

componentes CRISPR, mediante modificaciones en el plásmido y sirviéndose del método ensamblaje Golden Gate (Engler et al. 2009), éste puede llegar a albergar hasta siete sgRNAs en forma de DNA es decir, un vector de expresión todo en uno, con lo que se consigue una acción nucleasa dirigida a múltiples dianas simultáneamente, tantas como sgRNA distintos estén codificados en pX330 (Sakuma et al. 2014; Sakuma y Yamamoto 2015) (Figura 9).

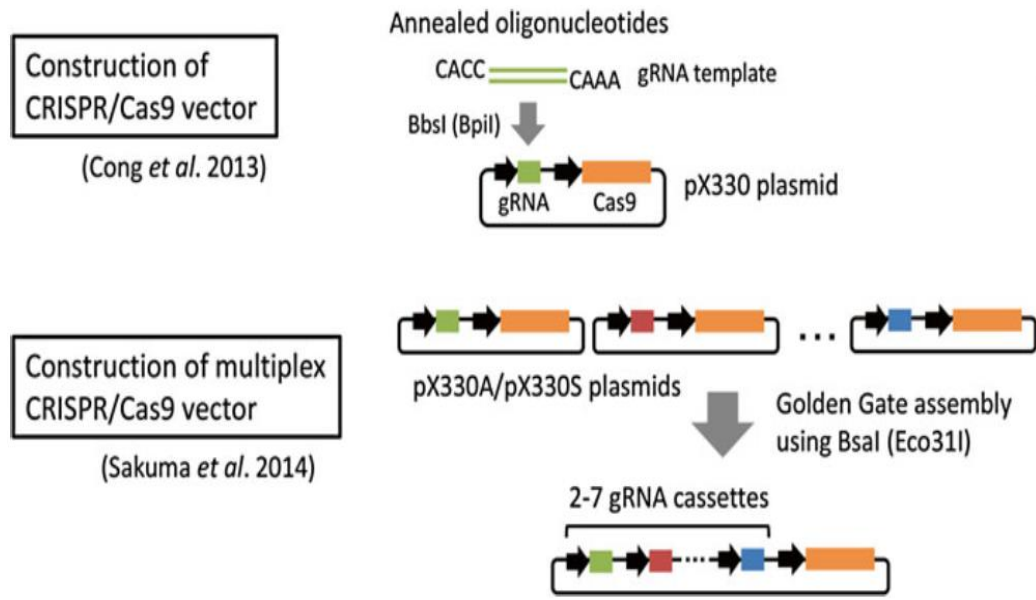


Figura 9 (Sakuma y Yamamoto 2015). Construcción de vectores de expresión para su introducción en las células a modificar. En la parte superior, vector con un solo sgRNA que podrá localizar una sola diana de corte. En la región inferior, vector con hasta siete sgRNAs capaz de guiar hacia tantas dianas como sgRNAs diferentes porte.

Tan importante como un buen diseño del sgRNA es la elección de un método propicio de introducción de la herramienta de edición genómica en base al tipo celular que se quiera modificar para conseguir de forma exitosa el cambio en el gen diana. En este sentido, se dispone de varios formatos de introducción de la nucleasa como son, en forma de DNA, mRNA o proteína siendo la transfección de DNA la elección más frecuente (Yamamoto 2015) (Figura 10). A continuación se exponen las distintas aproximaciones que se abordan para la transfección de la nucleasa y el sgRNA ya sea siguiendo métodos bioquímicos, físicos, virales o de fusión protéica.

En la transfección bioquímica del DNA portador de los componentes del sistema de edición genómica es decir, de vectores de expresión, el investigador se sirve de reactivos químicos de los que, previamente y de manera recomendable, se han

comparado eficiencias mediante su prueba en las células de estudio y con el vector de transfección control. Este método es altamente utilizado y popular gracias a su buena eficiencia y su propiedad de ser escalable y adaptable a distintas condiciones. Además, es el procedimiento más útil si se trabaja con células en cultivo o embriones.

Si se habla de métodos de transfección física, también se dispone de opciones variadas como son la electroporación, la nucleofección y la microinyección. Este tipo de mecanismos se escogen en el caso de tener células de estudio difíciles de transfectar como pueden ser las células troncales o primarias. Si el método elegido es la electroporación, estrategia mediante la cual se aplican pulsos eléctricos determinados por el tipo celular que crean poros transitorios en la membrana celular permitiendo que DNA, RNA y proteínas sean introducidos, lo más usual en relación a CRISPR-Cas9 es que se introduzca la Cas9 como proteína ya purificada junto con el sgRNA formando estructuras ribonucleoprotéicas de la maquinaria que conforma la herramienta ya ensamblada. Para facilitar la entrada de las ribonucleoproteínas al núcleo y aumentar así la eficiencia de la técnica, se desarrolló, posteriormente a la electroporación, la nucleofección, en la que además de aplicar pulsos eléctricos, se baña en una solución altamente amortiguadora de fuerza iónica. Siguiendo con las mejoras a la técnica inicial (la electroporación), a partir del siglo XXI se han desarrollado sistemas como el de transfección de neón que se sirve de un electrodo tipo pipeta en lugar de una cubeta o nuevos electroporadores cuyos pulsos eléctricos son de poración (corto y de alto voltaje) y de transferencia (largo y de bajo voltaje) para introducir la biopartícula de interés dentro de las células (Li y Hotta 2015). El último método físico a describir, la microinyección, consiste en introducir normalmente DNA (en el caso de CRISPR-Cas9 este DNA serían vectores de expresión) mediante una microaguja de alrededor de 10 μm y un micromanipulador en el plasma celular o núcleo (Sakuma y Yamamoto 2015).

Otra de las formas de transfección en relación a nucleasas específicas de diseño es la transfección mediante virus. Los vectores virales son herramientas tremendamente eficientes a la hora de introducir genes extraños dentro de células de mamífero dado que es ésta la función que realizan cuando infectan células en su estado natural. Hay descritos varios trabajos de edición genómica mediante CRISPR- Cas9 en los que se emplean vectores víricos de transfección tales como lentivirus o adenovirus (Abrahimi et al. 2015; Kabadi et al. 2014; Holkers et al. 2014). Además de en células de mamífero, en plantas, usualmente se emplea la infección mediada por *Agrobacterium* como método de inserción de CRISPR- Cas9 o, la transformación de protoplastos (Bortesi y Fischer 2014).

Por último, también se puede lograr la introducción de la herramienta a través de su conjugación con péptidos con capacidad penetrante a través de la membrana celular. En este caso la nucleasa Cas9 purificada ensamblada con el sgRNA pertinente formando ribonucleoproteínas se conjuga de modo químico con el péptido penetrante que actúa como guía para atravesar la membrana celular. Si esta fusión péptido-ribonucleoproteína simplemente se añade en el medio de cultivo, penetrará en las células, aunque también se puede facilitar su entrada mediante nucleofección (Li y Hotta 2015).

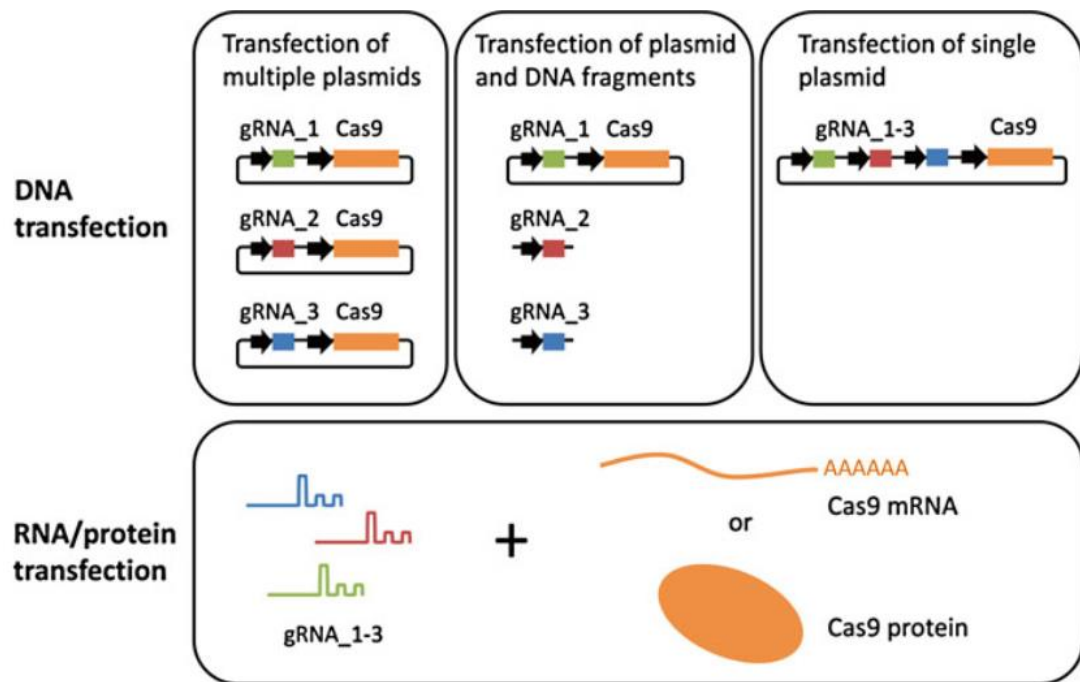


Figura 10 (Sakuma y Yamamoto 2015). Diferentes estrategias de transfección de CRISPR-Cas9 con múltiples dianas: introducción de múltiples vectores de expresión, de un solo plásmido y fragmentos de DNA codificantes para distintos sgRNAs, de un plásmido portador de varios sgRNAs y de transcritos tanto de sgRNA como de la Cas9 o de la Cas9 ya expresada en forma de proteína.

4.2.5 Inconvenientes del sistema CRISPR-Cas9: *Off target*

Si bien es cierto que en cuanto a la edición de genomas la herramienta CRISPR-Cas9 se perfila como una de las mejores tanto por su eficiencia como por sus rápidos resultados, fácil manejo y coste, existen ciertos matices que todavía tratan de optimizarse. La cuestión que más ocupa a científicos dedicados a la aplicación del sistema en edición genómica es la generación de *off target* o la obtención de modificaciones en dianas inespecíficas que no forman parte del locus objetivo. Estos

efectos *off target* se dan porque en el organismo de origen del sistema y en su función de defensa contra elementos nucleotídicos externos, puede ser útil el reconocimiento y destrucción de DNA viral o plasmídico hipervariable. Sin embargo, esta relativa flexibilidad en el reconocimiento de secuencias variables, cuando el sistema se saca de su contexto original y se dedica a la modificación genómica, genera mutaciones indeseadas más allá de la región diana y resulta en efectos colaterales indeseados (Ma et al. 2014). Los efectos *off target* varían mucho dependiendo de dónde se encuentre el fallo en el apareamiento o *mismatch* y el número de ellos, por ejemplo, según estudios, si el *mismatch* se encuentra en el extremo 5' de la diana pero no entre 8 y 12 pares de bases aguas arriba de la secuencia PAM, es tolerado (Cong et al. 2013) en cambio, si hay tres *mismatches* y están fuera del extremo 5' o éstos se encuentran en una cantidad de hasta cinco nucleótidos, sí se producen eventos de mutaciones *off-target* pudiendo llegar éstas a tasas iguales a las de modificaciones en la diana correcta (Fu et al. 2013). Además de esto, es lícito pensar que las diferencias entre efectos *off-target* en determinado tipo celular dependen de la epigenómica y la estructura de la cromatina (Ma et al. 2014). Para evitar los efectos indeseados de los *off-target* se hace necesario incidir en el diseño del RNA guía evitando que éste pudiera coincidir con más de una secuencia tanto codificante como no codificante mediante el estudio de la secuenciación del genoma de la célula a modificar y sirviéndose de herramientas de búsqueda de *off-target* disponibles en páginas de internet especializadas en CRISPR-Cas. El problema de la generación de *off-target* se ha abordado a través de múltiples estrategias como son:

- La ya mencionada anteriormente buena selección de la diana junto con un buen diseño del sgRNA. A este respecto, algunos estudios verifican que un alto contenido en G+C promueve una mejor hibridación mientras que un bajo contenido en estas bases así como un bucle de desapareamiento en la unión DNA-RNA aumentan los *off-target* (Lin et al. 2014). En definitiva, se debe tratar, en la medida de lo posible, de aparear las bases del sgRNA y la diana de modo que éstas coincidan al 100%.
- Una disminución en la cantidad de enzima y sgRNA que se introduce, esto es, una disminución en la concentración del complejo Cas9-sgRNA en el plásmido de expresión, disminuye los *off-target* aunque, todo sea dicho, también hace que decrezcan los *on-target* o unión a la diana específica (Hsu et al. 2013).
- La introducción en forma de proteína purificada de Cas9 y su correspondiente sgRNA de diseño con el fin de tener la enzima en una concentración determinada que no vaya en aumento como podría ocurrir si se

introducen ambos componentes en un plásmido que continuase expresándolos y así aumentando su unión a dianas inespecíficas (Kim y Kim 2014).

- El acortamiento del sgRNA en su extremo 3' y la adición de dos GG a continuación de la región complementaria al DNA llegando a discriminar mediante estas modificaciones, entre una diana correcta y una incorrecta que difieren solamente en dos bases (Cho et al. 2014).
- El empleo, como ya se ha mencionado en párrafos anteriores de nickasas.
- El acortamiento del sgRNA en su extremo 5' que trunca la secuencia complementaria en un tamaño de 17 o 18 pares de bases en lugar de 20 hecho que incrementa su especificidad e inflexibilidad (Fu et al. 2014).
- La fusión de nucleasas Fok I a sendas enzimas Cas9 con sus dominios catalíticos inactivos y sus respectivos RNAs guía, este conjunto es nombrado como fCas9. La fusión a Fok I obliga a la dimerización para que se produzca el corte entre dos sgRNAs separados por entre 15 y 25 pares de bases y con sus PAMs orientados hacia fuera (Figura 11). Es esta dimerización la que otorga a este sistema cuatro veces más especificidad (medida como el ratio *on-target* / *off-target*) que el empleo de nickasas y una especificidad más de 140 veces mayor que la de Cas9 silvestre, prevaleciendo la especificidad por encima de la eficiencia de la enzima WT en este tipo de construcciones pero siendo similar a la eficiencia de las nickasas (Tsai et al. 2014; Guilinger et al. 2014).

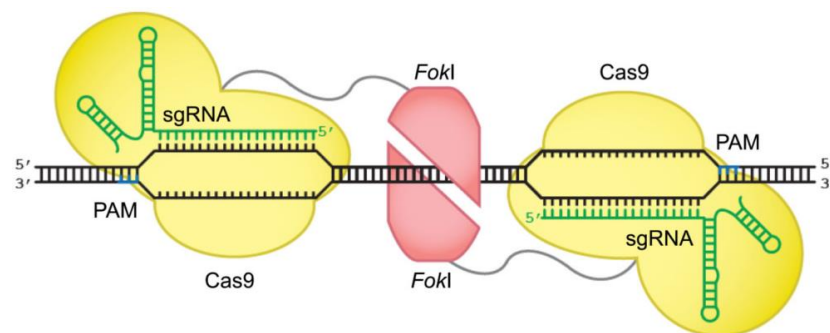


Figura 11 (Guilinger et al. 2014). Fusión de complejos Cas9-sgRNA sin actividad catalítica a nucleasas Fok I formando conjuntos fCas9. Se observa la obligatoria dimerización de los complejos para generar el DSB.

4.2.6 Perspectivas de aplicaciones potenciales

Como herramienta aplicada a la manipulación genómica, CRISPR-Cas9 tiene todas las posibilidades de explotación que la manipulación genómica puede albergar. En este apartado se citarán las que se consideran más prometedoras como pueden ser, a grandes rasgos, la biomedicina y terapia génica, la biotecnología en su más amplio espectro y la investigación.

En cuanto a la aplicación terapéutica de Cas9, son muchos los investigadores que tienen puestas sus miradas en la optimización de la herramienta para poder tratar diferentes tipos de desórdenes. En patologías que implican desórdenes genéticos o factores genéticos de riesgo se recurriría al mecanismo de reparación inducido por Cas9 que permitiría, idealmente, revertir la secuencia patológica en una secuencia silvestre. En el caso de encontrarse frente a genomas invasivos patogénicos o mutaciones dominantes negativas, se abordaría su tratamiento mediante la disrupción génica a la que puede dar lugar la herramienta. Además de esto, no hay que olvidar que Cas9 puede actuar, con las fusiones pertinentes, como activador o represor transcripcional, algo que permitiría, por ejemplo, reprimir la transcripción de oncogenes y receptores virales en células hospedadoras o, por el contrario, activar transcripcionalmente a los genes supresores de tumores o a las globinas. Todos estos usos, podrían llevar a corregir desregulaciones epigenéticas, controlar los procesos inflamatorios y la autoinmunidad y reprimir la transcripción de genes virales o correceptores virales en tipos celulares vulnerables (Mali, Esvelt, et al. 2013). Hay que recalcar que, aunque aún haya que implementar la técnica en cuanto a los *off target*, la vehiculización y la posible respuesta inmune, ya ha habido estudios con resultados exitosos de corrección de mutaciones en ratón tanto *in vivo* en ratones adultos (Yin et al. 2014) como *in vitro* en línea germinal (Long et al. 2014) y de eliminación selectiva de las bacterias virulentas de una comunidad mediante fagos portadores de Cas9 que tenían como diana a los genes virulentos de dichas bacterias (Bikard et al. 2014).

A nivel de la aplicación de la herramienta CRISPR-Cas9 de edición genómica en el ámbito biotecnológico y entendiendo la biotecnología como toda aplicación tecnológica que utilice sistemas biológicos y organismos vivos o sus derivados para la creación o modificación de productos o procesos para usos específicos (United Nations 1992) aparte de en lo referente a la biomedicina y el tratamiento de patologías, es innegable el valor del sistema como herramienta para implementar, mediante la modificación genómica, rutas en biotecnología industrial que permitan, de uno u otro modo, aumentar el rendimiento de procesos como pueden ser la producción de combustibles o de

biomateriales. Siendo también clave su empleo en sectores como la agricultura, mejorando las características de ciertos cultivos (Wang et al. 2014) o controlando su proliferación, la farmacia, mejorando rutas de producción o el desarrollo de fármacos o cosméticos provenientes de organismos biológicos, o la ciencia de los alimentos.

Dentro del ámbito de la investigación biológica, el sistema CRISPR-Cas9 mediante las variadas formas en las que es capaz de modificar el genoma (reparación, *knockout*, regulación de la expresión), se puede aplicar a estudios de identificación y función de genes silvestres en animales y sistemas celulares modelo así como a la identificación de mutaciones genéticas o variantes epigenéticas asociadas a alteraciones de funciones biológicas o fenotipos patológicos. Hay un sinnúmero de usos otorgables a la herramienta dentro del área de la investigación biológica tanto por su capacidad de corte en dianas específicas como por su capacidad, tras eliminar su competencia nucleasa, señalizadora en esas mismas dianas.

4.3 Controversia despertada: patentes e implicaciones éticas

El mecanismo de edición de genomas mediante la Cas9, como muchos de los sistemas capaces de alterar de forma dirigida y concreta los genomas de distintos tipos de células, no está exento de cierta polémica. Esta controversia básicamente radica en la inquietud que genera el hecho de poder reescribir los genomas de líneas germinales y, de manera más principal y concreta, de embriones humanos. El temor se fundamenta en la facilidad de aplicación del método de edición que podría dar lugar a un uso irresponsable de éste en, por ejemplo, clínicas de fertilidad para no sólo poder revertir mutaciones patológicas sino para crear embriones “a la carta” con unas características físicas específicas. Además de esto, al no ser un método todavía depurado al 100%, dentro de la comunidad científica, también preocupa su aplicación en embriones humanos y pacientes con enfermedades genéticas por generar mutaciones en lugares no diana u *off target* que, aparte de afectar al individuo al que daría lugar dicho embrión o al paciente, afectarían, también, a sus posibles descendientes teniendo en cuenta que se trata de mutaciones heredables. A este respecto, la bioquímica Jennifer Doudna afirma que “sería necesario decidir, para cada potencial aplicación, si los riesgos son mayores que las posibilidades de beneficio para un paciente. Creo que esta valoración debe ser hecha caso por caso” (Cyranoski 2015). Por otro lado, la legislación que ampara o prohíbe este tipo de procedimientos de edición genómica en líneas germinales es muy distinta entre países (Araki y Ishii 2014) haciéndose difícil un control sobre su uso.

Dentro del marco de las implicaciones éticas, se hace indispensable citar la publicación en abril de este año de un artículo rechazado por las revistas *Nature* y *Science* y, finalmente publicado en la revista *Protein & Cell*, en el cuál Junjiu Huang y sus colegas exponen su trabajo en la edición genómica de cigotos humanos tripnucleares no viables con el objetivo de probar, por vez primera, la técnica en embriones humanos y su eficacia en la corrección del gen responsable de la β -talasemia (Liang et al. 2015). Huang mantiene que, los embriones humanos no viables empleados en su estudio constituyen un modelo más cercano a embriones humanos normales y asevera que “queríamos mostrar nuestros datos al mundo para que la gente sepa qué es lo que pasa realmente con este modelo en lugar de solamente hablar de lo que podría pasar sin datos” (Cyranoski y Reardon 2015). Mientras que Edward Lanphier comunica que “esto (refiriéndose a la publicación de Huang) subraya lo que decíamos antes: necesitamos pausar esta investigación y estar seguros de que tenemos una discusión amplia sobre hacia qué dirección estamos yendo” (Lanphier et al. 2015).

Otro aspecto que despierta cierto debate, es el tema de las patentes. En fecha de 8 de julio de 2014, F. Zhang como inventor y el *Broad Institute Inc.* junto con el *Massachusetts Institute of Technology* ambos de Cambridge como solicitantes, obtuvieron la patente estadounidense de los sistemas CRISPR-Cas y sus métodos para alterar la expresión de productos génicos (Zhang 2014). Entretanto, también solicitaron la correspondiente patente europea publicada como EP2764103 A0 y aún sin resolver. Sin embargo, tras ser solicitada esta última, entró en proceso de prosecución en la Oficina Europea de Patentes (EPO) (Kupecz 2014). Esto sucede porque cuando se hace una solicitud de patente en la EPO, se procede a su examinación y pueden tomar parte terceras personas que, de manera anónima, presentan sus objeciones a la patente. En este caso, estas terceras partes, han aportado más de 20 referencias de artículos en contra de la patente de Zhang de forma previa al estudio de ésta en la División de Examen de Patentes. Estos datos aportados, tratan de demostrar y prevenir la concesión de la patente a Zhang basándose en su falta de novedad e inventiva (Kupecz 2014). No obstante, no es descabellado pensar que estas terceras partes podrían ser empresas biotecnológicas con el objetivo de poseer una patente que, a todas luces y, teniendo en cuenta el auge de las aplicaciones de CRISPR-Cas, reportaría grandes beneficios. Por tanto, se podría decir, que la problemática de la patente europea, responde al conflicto entre entidades que desean los derechos de explotación del sistema. Así pues, en caso de que la patente europea, finalmente se otorgue, sería diferente en varios puntos respecto de la estadounidense (Figura 12).

| Table 1 The Zhang patent: comparison of US granted claims and current EP claims (as of 17 November 2014). | |
|---|---|
| EP claim 1 | US claim 1 |
| <p>1. An engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) vector system comprising one or more vectors comprising:</p> <p>a) a first regulatory element operably linked to one or more nucleotide sequences encoding one or more CRISPR-Cas system guide RNAs that hybridize with target sequences in polynucleotide loci in a eukaryotic cell,</p> <p>b) a second regulatory element operably linked to a nucleotide sequence encoding a Type II Cas9 protein,</p> <p>wherein components (a) and (b) are located on same or different vectors of the system,</p> <p>wherein the guide RNA comprises a tracr sequence which is 30 or more nucleotides in length</p> <p>whereby the one or more guide RNAs target the polynucleotide loci in a eukaryotic cell and the Cas9 protein cleaves the polynucleotide loci, whereby sequence of the polynucleotide loci is modified; and, wherein the Cas9 protein and the one or more guide RNAs do not naturally occur together.</p> | <p>1. A method of altering expression of at least one gene product comprising introducing into a eukaryotic cell containing and expressing a DNA molecule having a target sequence and encoding the gene product an engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)—CRISPR associated (Cas) (CRISPR-Cas) system comprising one or more vectors comprising:</p> <p>a) a first regulatory element operable in a eukaryotic cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence, and</p> <p>b) a second regulatory element operable in a eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein,</p> <p>wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNA targets the target sequence and the Cas9 protein cleaves the DNA molecule, whereby expression of at least one gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.</p> |
| The applicant is free to amend its claims throughout the examination process at the EPO, so further claim sets may be submitted. The most recent claims can be found in the EPO register ⁶ . | |

Figura 12 (Kupecz 2014). Diferencias entre las peticiones o demandas de la patente europea y estadounidense.

4.4 Simulación práctica: sgRNA

Llegados a este punto, y como parte del presente Trabajo de fin de Grado, se propone una parte práctica en la que se pretende simular, mediante programas informáticos, el diseño de un sgRNA. Para esto, el modo de acción o mecanismo a desarrollar por la Cas9 que se ha elegido es el de reparación. Se escoge, como gen a reparar, el responsable de una enfermedad monogenética clásica como es la anemia falciforme ya que no se registran, hasta la fecha, artículos que apliquen la reparación genómica para tratar esta enfermedad mediante la herramienta descrita a lo largo de las páginas anteriores.

4.4.1 La enfermedad

La anemia falciforme es una enfermedad monogenética causada por una mutación de tipo SNP en el gen codificante para la β -globina denominada de forma abreviada HbS (de hemoglobina S). El gen, se localiza en el cromosoma 11 en la posición 5227002 (según el *Genome Reference Consortium Human Build 38* o GRCh38). El cambio se produce en la posición 17 de la secuencia nucleotídica correspondiente al gen donde se cambia una adenina por una timina, hecho que da lugar a la aparición del aminoácido valina en lugar de glutamato. La consecuencia, a nivel de estructura proteico, de la mutación, es la generación de un motivo hidrofóbico entre las dos moléculas de β -globina presentes en el tetrámero desoxigenado de la hemoglobina (formado por dos unidades de α -globina y dos de β -globina). Así, en esta enfermedad, ambas subunidades de β -globina de la hemoglobina son HbS (Figura 13). El efecto de esta conformación es la disrupción de la arquitectura (pasa a ser falciforme) y flexibilidad del

eritrocito, la temprana deshidratación celular que conlleva daño oxidativo y, finalmente, la hemólisis (Rees et al. 2010).

Los síntomas, siendo más o menos severos, comienzan en la infancia con un cuadro de bajo recuento de células sanguíneas de línea roja (hematíes), infecciones recurrentes y episodios de dolor por atasco de eritrocitos falciformes en vasos sanguíneos pequeños. Las consecuencias de la anemia pueden ser dificultades respiratorias, fatiga, y retraso en el crecimiento y desarrollo en la niñez. Mientras que la hemólisis puede causar amarilleamiento de ojos y piel, signo de ictericia. El atasco de glóbulos rojos en los vasos sanguíneos, a su vez, genera insuficiencia de oxígeno en tejidos y órganos como pulmones, riñones, bazo y cerebro. Una de las complicaciones severas de la enfermedad es la hipertensión pulmonar que puede dar lugar a insuficiencia cardíaca y muerte (Figura 13) (Genetics Home Reference, 2015).

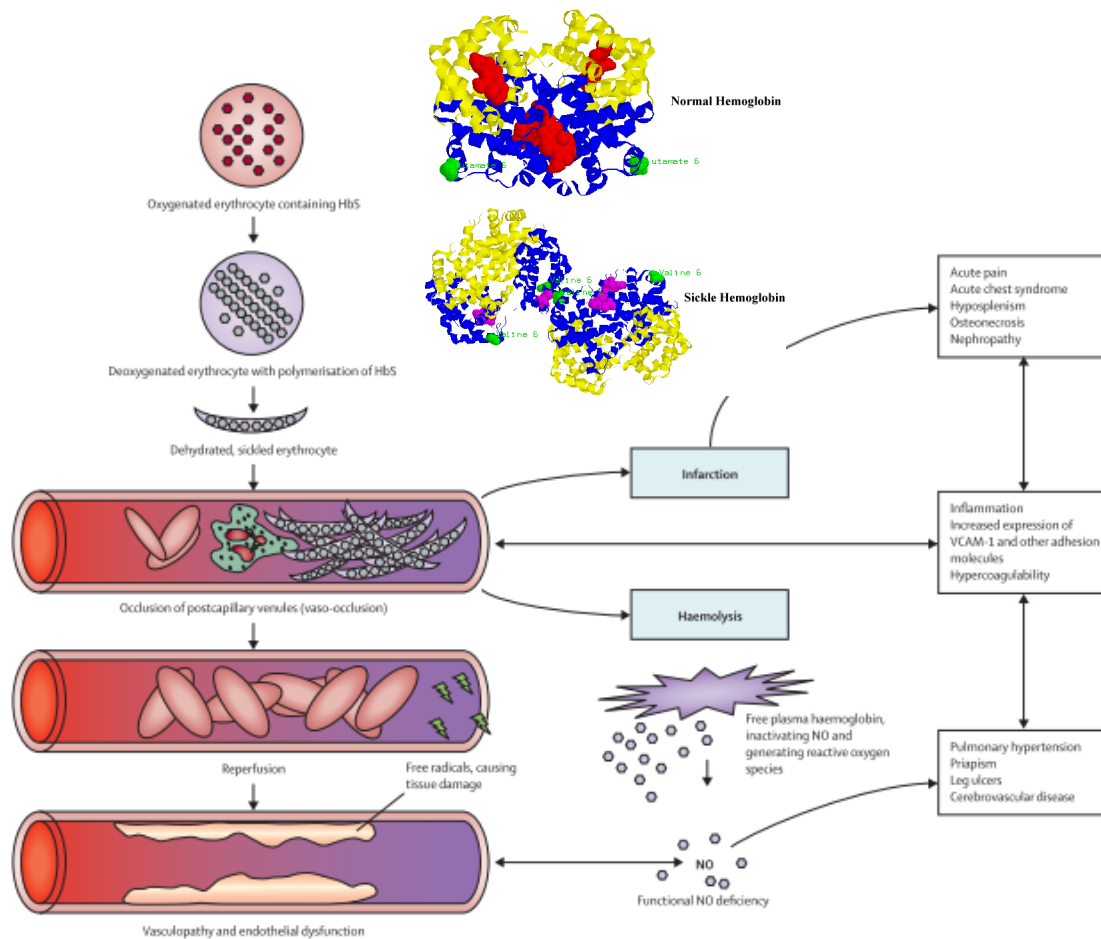


Figura 13 (Rees et al. 2010). En la parte superior de la imagen se aprecian las diferencias conformacionales entre una molécula de hemoglobina silvestre y una patológica. El esquema ilustra los síntomas y complicaciones que pueden sufrir los pacientes de anemia falciforme.

4.4.2 Diseño del sgRNA

Para comenzar con la simulación de los pasos a seguir en un hipotético estudio en el que se intente corregir la mutación del gen HBB mediante el uso de CRISPR-Cas9, primero, se debe encontrar la secuencia codificante para el SNP a reparar. Para ello, se busca en la bibliografía disponible el código rs (número otorgado al SNP dentro de la base de datos dbSNP) correspondiente a la mutación HbS que, en este caso, es el rs334 (Steinberg y Sebastiani 2012). Una vez se conoce el código rs, se accede a la base de datos dbSNP (dbSNP, 2015) y se inserta el código rs como criterio de búsqueda. Mediante este motor de búsqueda con el código rs334 introducido, se obtienen 4 resultados entre los cuales se selecciona y abre el primero (en el que se especifica el código rs334 y la especie *Homo sapiens*). Con esta referencia abierta, se observa que, efectivamente, el código rs334 corresponde a la mutación que da lugar a subunidades HbS patológicas por un cambio del nucleótido adenina por timina (Figura 14 A). A partir de esta página, y clicando en la opción Fasta, se extrae y copia la secuencia en dicho formato (Figura 14 B).

A

Reference SNP (refSNP) Cluster Report: rs334 ** With Pathogenic, other allele **

| RefSNP | Allele | HGVS Names |
|---|--|---------------------------|
| Organism: human (<i>Homo sapiens</i>) | Variation Class: SNV | NC_000011.10:g.5227002T>A |
| Molecule Type: Genomic | single nucleotide variation | NC_000011.10:g.5227002T>C |
| Created/Updated in build: 36/144 | RefSNP Alleles: A/C/G/T (REV) | NC_000011.9:g.5248232T>A |
| Map to Genome Build: 107/144 | A: germline | NC_000011.9:g.5248232T>C |
| Validation Status: >= 100 | C: germline | NC_000011.9:g.5248232T>G |
| Citation: PubMed | G: germline | NC_000007.3:g.706144A>C |
| | T: germline | NC_000007.3:g.706144A>G |
| | Ancestral Allele: A | NC_000007.3:g.706144A>T |
| | Variation Viewer: View | NM_000518.4:c.20A>G |
| | Clinical Significance: With Pathogenic, other allele | NM_000518.4:c.20A>G |
| | MAF/MinorAlleleCount: A=0.0274/137 | |
| | MAF Source: 1000 Genomes | |

SNP Details are organized in the following sections:
[GeneView](#) [Map](#) [Submission](#) [Fasta](#) [Resource](#) [Diversity](#) [Validation](#)

B

Fasta sequence (Legend)

```

>gn1|dbSNP|rs334|allelePos=301|totalLen=601|taxid=9606|snpclass=1|alleles='A/C/G/T'|mol=Genomic|build=144
GCAATTTGTA CTGATGGTAT GGGGCAAGA GATATATCTT AGAGGAGAGG CTGAGGGTTT
GAAGTCCAC TCCTAAGCA GTGCGAGAG AGCAAGGAC AGGTACGCT GTCATGACTT
AGACCTCACC CTGTGAGGCC ACACCTAGG GTTGCCCAAT CTACTCCAG GAGCAGGAG
GGCAGAGGCC AGGCTGGG ATAAAGTCA GGGCAGAGCC ATCTATTGCT TACATTGCT
TCTGACACAA CTGTGTTTAC TAGCAACCTC AACAGAGAC CATGGTGCAT CTGACTCCTG
N
GGAGAAGTGT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGATGAAG TTGGTGGTGA
GGCCTGGGCG AGGTTGGTAT CAAGGTTACA AGACAGGTTT AAGAGACCA ATAGAAACTG
GGCATGTGGA GACAGAGAG ACTCTTGGGT TTCTGATAGG CACTGACTCT CTCTGCTAT
TGGTCTATT TCCACACCTT AGGCTGCTGG TGGTCTACCC TTGAGCCAG AGGTTCTTTG
AGTCTTTGG GATCTGTGCT ACTCTGTGAT CTGTATGGG CAACCTTAGG GTGAAGGCTC

```

Figura (dbSNP, 2015). A- Página correspondiente a la referencia rs334. En rojo, en la zona central, se observa la indicación de la patogenicidad del alelo. B- Secuencia en formato Fasta. La N solitaria en color verde corresponde al nucleótido mutado, en este caso, a T.

Tras obtener la secuencia en formato Fasta, se procede al diseño del sgRNA que señalice la diana de corte para Cas9. En este punto, se hace uso de la herramienta E-CRISP (Heigwer et al. 2014) creada por el laboratorio de Boutros englobado dentro del Centro de Investigación del Cáncer de Alemania o dkfz. Se elige esta herramienta frente a otras como puede ser *Optimized CRISPR Design* sita en el laboratorio de Zhang, localizado en el MIT, por sus cualidades *user friendly* de uso e interpretación de datos y resultados. Para poner en marcha el diseño del sgRNA, una vez dentro del portal web de la herramienta E-CRISP, se selecciona el apartado *De-novo* (Figura 15 A). En la nueva pantalla que aparece, se elige la opción *Homo sapiens sapiens* (Human, GRCh38.77), se escoge la alternativa Enter Target Sequence, se introducen 121 bases de la secuencia codificante para HbS conteniendo la mutación (en este caso, se cambia la N en color verde por una T), se deja el parámetro *medium* que viene predeterminado en la página y, por último, se hace clic en *Start sgRNA search* (Figura 15 B). En este punto, es necesario aclarar que, al tratarse de una simulación de prueba o ejemplo, no se han tenido en cuenta las opciones avanzadas que ofrece la herramienta como son, entre muchas otras, la elección del propósito del sgRNA dentro del experimento (disrupción del gen, activación etc.), la longitud del sgRNA tras la secuencia PAM, el porcentaje de cada uno de los nucleótidos o la anotación de los genes filtrados. A partir de esta serie de movimientos, aparece una nueva pantalla con los resultados. Se debe seleccionar como sgRNA el resultado que más cerca hibride de la mutación de interés y que, a su vez, tenga los parámetros de eficiencia y especificidad más altos, sin olvidar el número de dianas. En este caso, se elegiría el sgRNA HbS_5_80 localizado en tercer lugar en la tabla (Figura 15 C) atendiendo a los citados parámetros. Además, si se desliza hacia abajo la página, se observa un esquema en el que aparece la secuencia insertada en azul y, en verde, los diferentes resultados de sgRNA (Figura 15 D). Es en esta ilustración dónde se comprueba que, efectivamente, es el HbS_5_80 el sgRNA el que cumple las mejores condiciones de cercanía a la mutación (a 21 bases de distancia) que se localizaría en la posición 61 de esta secuencia. Igualmente, el corte del DNA por parte de la Cas9, se generaría a menos de 50 pares de bases de la mutación, con lo que el diseño del molde de reparación sería en forma de ssODN y, por tanto, esto atiende al diseño más eficiente.

Posteriormente a la obtención de un sgRNA adecuado, se debería valorar más a fondo la existencia de *off target* en el genoma completo para el mismo, así como la puesta en marcha de todo el protocolo necesario para realizar un estudio de reparación de mutación tipo SNP mediante la herramienta de ingeniería genómica CRISPR-Cas9.

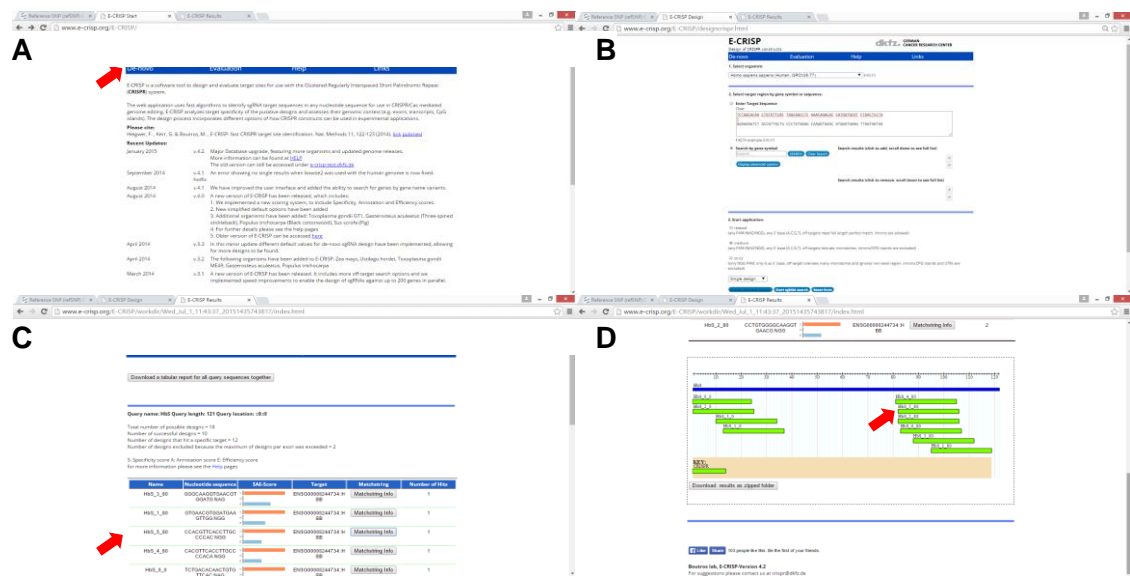


Figura (Heigwer et al. 2014). A- Página de inicio del servidor que aloja la herramienta E-CRISP. La flecha roja indica la sección en la que debemos clicar para que aparezca la herramienta. B- Pantalla dónde se especifica la especie para la cual se requiere el sgRNA, se inserta la secuencia y se deciden los parámetros de diseño. C y D- Página de resultados con los diez diseños que propone como exitosos el programa. En la tabla se detallan tanto el nombre del sgRNA como su secuencia, su especificidad y eficiencia, su diana y el número de éstas. En el esquema se ilustra la posición de la diana de los sgRNAs diseñados en relación a la secuencia génica insertada en el programa. Las flechas rojas, en ambos casos, señalan el diseño HbS_5_80, la opción mejor considerada teniendo en cuenta todos los criterios de selección aplicados.

5. PLAN DE TRABAJO

Al tratarse de un Trabajo de Fin de Grado (TFG) esencialmente bibliográfico, el plan de trabajo adoptado, no comporta tantos apartados como los que acontecerían en un estudio de carácter práctico. Asimismo los límites de las tareas desarrolladas son bastante difusos en el tiempo dando lugar al solapamiento de unas con otras. Con todo, se considera la siguiente división de actividades:

- Asignación definitiva del TFG.
- Familiarización con la herramienta de gestión de bibliografía Mendeley Desktop (Mendeley, 2015).
- Lectura de artículos científicos proporcionados por el tutor.
- Elaboración de la estructura del trabajo a seguir.

- Búsqueda bibliográfica de cada uno de los apartados.
- Redacción y elaboración del TFG incluída la parte práctica.
- Correcciones por parte del tutor.
- Entrega.

| | | | | | | | | | |
|-----------------------------------|----------------|--|--|--|--|--|--|--|--|
| Asignación | Noviembre 2015 | | | | | | | | |
| Familiarización con Mendeley | Diciembre 2015 | | | | | | | | |
| Lectura de artículos del tutor | Enero 2015 | | | | | | | | |
| Diseño de estructura a seguir | Febrero 2015 | | | | | | | | |
| Búsqueda bibliográfica específica | Marzo 2015 | | | | | | | | |
| Redacción | Abril 2015 | | | | | | | | |
| Correcciones del tutor | Mayo 2015 | | | | | | | | |
| Entrega | Junio 2015 | | | | | | | | |
| | Julio 2015 | | | | | | | | |

6. CONCLUSIONES

Habiendo revisado todos los puntos considerados clave en relación a la aplicación de CRISPR-Cas como sistema que permite editar genomas, se llega a una serie de conclusiones. Entre ellas destaca la importancia de la investigación básica, gracias a la cual, a lo largo del tiempo, el descubrimiento de microorganismos portadores de unas secuencias a las que no se conocía función ha permitido, en la actualidad, la creación y el uso de una de las herramientas más poderosas, asequibles tanto en precio como en facilidad de uso, versátiles y prometedoras para la ingeniería biológica. Además de esto, se ha de remarcar la aplicabilidad potencial de la herramienta en cantidad de procesos y áreas de la ciencia como mejora de procesos industriales, medicina y farmacología, mejora de cultivos y un largo etcétera. Si bien es cierto que aún queda un largo recorrido en cuanto a la optimización de sus resultados y las implicaciones éticas que supone la aplicación de todo sistema de edición de genes, se cree probable que, en un futuro, la ingeniería genómica sea una solución factible a grandes problemas de la humanidad.

Having reviewed all the items considered as the key related to the application of CRISPR-Cas as a system that enables editing genomes, a range of conclusions are reached. Including most notably the importance of basic research, thanks to which, over time, the discovery of microorganisms carrying sequences with unknown function has allowed, nowadays, the creation and use of one of the most powerful, affordable in price and ease-of-use, versatile and promising biological engineering tool. In addition to this, it must be highlighted the tool's potential applicability in many processes and science areas like improving industrial processes, medicine and pharmacology, crop improvement and so on. While there is still a long way concerning to the optimization of its results and the ethical implications involving the implementation of any editing gene system, is likely that, in the future, genetic engineering will be a feasible solution for the major problems of humankind.

7. BIBLIOGRAFÍA

Abrahimi, P. et al., 2015. Efficient Gene Disruption in Cultured Primary Human Endothelial Cells by CRISPR/Cas9. *Circulation Research*, (April 2015) pp1-13

Addgene, CRISPR in the Lab: A Practical Guide. Disponible en: <http://www.addgene.org/crispr/guide/>. Consultada el 15 de abril de 2015.

Araki, M. e Ishii, T., 2014. International regulatory landscape and integration of corrective genome editing into in vitro fertilization. *Reproductive Biology and Endocrinology*, 12(108), pp.1–12.

Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N. Y.)*, 315(5819), pp.1709–1712.

Barrangou, R. y Marraffini, L.A., 2014. CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell*, 54(2), pp.234–244.

Barrangou, R. y van der Oost, J., 2013. *CRISPR-Cas Systems RNA-Mediated Adaptive Immunity in Bacteria and Archaea*, Springer.

Bhaya, D., Davison, M. y Barrangou, R., 2011. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics*, 45, pp.273–297.

Bikard, D. et al., 2014. Development of sequence-specific antimicrobials based on programmable CRISPR-Cas nucleases. *Nature biotechnology*, 32(11), pp.1146–1150.

Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151(8), pp.2551–2561.

Bortesi, L. y Fischer, R., 2014. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, 33(1), pp.41–52.

Brouns, S.J.J. et al., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N. Y.)*, 321(5891), pp.960–964.

Chavez, A. et al., 2015. Highly efficient Cas9-mediated transcriptional programming. *Nature Methods*, (February), pp.1–5. Disponible en: <http://dx.doi.org/10.1038/nmeth.3312>.

Cho, S.W. et al., 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31(3), pp.230–2.

Cho, S.W. et al., 2014. Analysis of off-target effects of CRISPR Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, 24, pp.132–141.

Chylinski, K., Le Rhun, A. & Charpentier, E., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology*, 10(5), pp.726–37.

- Cong, L. & Zhang, F., 2015. Genome Engineering Using CRISPR-Cas9 System. En: Chromosomal Mutagenesis, Methods in Molecular Biology. Pruet-Miller, S.M. Springer, 1239, pp 197-217.
- Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339(6121), pp.819–823.
- Cox, D.B.T., Platt, R.J. y Zhang, F., 2015. Therapeutic genome editing: prospects and challenges. *Nature Medicine*, 21(2), pp.121–131.
- Cyranoski, D., 2015. Scientists sound alarm over DNA editing of human embryos. *Nature news*, pp. 1-3.
- Cyranoski, D. y Reardon, S., 2015. Chinese scientists genetically modify human embryos. *Nature news*, pp.1–9.
- Davis, E., 2014. CRISPR- Cas9 Specificity: Taming Off- target Mutagenesis. *Genecopoeia* , pp.1–4.
- dbSNP. Database of Single Nucleotide Polymorphisms (dbSNP). National Center for Biotechnology Information, U.S. National Library of Medicine (Bethesda, MD, US). <http://www.ncbi.nlm.nih.gov/SNP/>
- Deltcheva, E. et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602–607.
- Dicarlo, J.E. et al., 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, 41, pp.4336–4343.
- Doudna, J.A. y Charpentier, E., 2014. Genome Editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213).
- Engler, C. et al., 2009. Golden gate shuffling: A one-pot DNA shuffling method based on type IIS restriction enzymes. *PLoS ONE*, 4(5), pp. 1-9.
- Friedland, A.E. et al., 2013. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature methods*, 10(8), pp.741–3.
- Fu, Y. et al., 2013. High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, 31(9), pp.822–826.
- Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(3), pp.279–84.
- Garneau, J.E. et al., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), pp.67–71.
- Gasiunas, G. et al., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, 109(39), pp.E2579–E2586.

Genetics Home Reference. National Center for Biotechnology Information, U.S. National Library of Medicine (Bethesda, MD, US).

<http://ghr.nlm.nih.gov/condition/sickle-cell-disease/>

Gilbert, L.A. et al., 2014. Resource Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 159, pp.647–661.

Guilinger, J.P., Thompson, D.B. & Liu, D.R., 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology*, 32(6), pp.577–82.

Hale, C.R. et al., 2009. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, 139(5), pp.945–956.

Heigwer, F., Kerr, G. y Boutros, M., 2014. E-CRISP: fast CRISPR target site identification. *Nature methods*, 11(2), pp.122–123.

Holkers, M. et al., 2014. Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nature methods*, (august), pp. 1-7.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–32.

Hsu, P.D., Lander, E.S. y Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157, pp.1262–1278.

Hwang, W.Y. et al., 2013. Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS ONE*, 8(7), pp.1–9.

Ishino, Y. et al., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429–5433.

Jackson, D. a, Symons, R.H. & Berg, P., 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 69(10), pp.2904–2909.

Jansen, R. et al., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), pp.1565–1575.

Jiang, W. et al., 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Research*, 41(20).

Jiang, W. et al., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, 31(3), pp.233–9. Jinek, M. et al., 2012. A Programmable Dual-RNA – Guided. *Science*, 337(August), pp.816–822.

Jinek, M. et al., 2013. RNA-programmed genome editing in human cells. *eLife*, 2013(2), pp.1–9.

- Jinek, M. et al., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, 343(6176), p.1215-1226.
- Kabadi, A.M. et al., 2014. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Research*, 42(19), p.gku749.
- Kim, H. y Kim, J.-S., 2014. A guide to genome engineering with programmable nucleases. *Nature reviews. Genetics*, 15(5), pp.321–34.
- Konermann, S. et al., 2014. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536), pp.583–588.
- Kupecz, A., 2014. Who owns CRISPR-Cas9 in Europe? *Nature Biotechnology*, 32(12), pp.1194–1196.
- Lanphier, E. et al., 2015. Don't edit the human germ line Heritable. *Nature*, 519, pp.410-411.
- Larson, M.H. et al., 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols*, 8(11), pp.2180–2196.
- Li, H.L. y Hotta, A., 2015. Editing Cultured Human Cells: From Cell Lines to iPS Cells. En *Targeted Genome Editing Using Site-Specific Nucleases*. T. Yamamoto, ed. Springer, p. 205.
- Liang, P. et al., 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell*, 6(5), pp.363–372.
- Lin, Y. et al., 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research*, 42(11), pp.7473–7485.
- Liu, L. y Fan, X.D., 2014. CRISPR-Cas system: A powerful tool for genome engineering. *Plant Molecular Biology*, 85, pp.209–218.
- Long, C. et al., 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science*, 345(6201), pp.1184–1188.
- Ma, S. et al., 2014. CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. *Scientific reports*, 4, p.4489.
- Ma, Y., Zhang, L. y Huang, X., 2014. Genome modification by CRISPR / Cas9. *FEBS Journal*, 281, pp.5186–5193.
- Maeder, M.L. et al., 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature methods*, 10, pp.977–979.
- Makarova, K.S., Brouns, S.J.J., et al., 2011. Evolution and classification of the CRISPR-Cas Systems. *Nature reviews. Microbiology*, 9(6), pp.467–477.
- Makarova, K.S., Aravind, L., et al., 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct*, 6(1), p.38.

- Mali, P., Yang, L., et al., 2013. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–6.
- Mali, P., Esvelt, K.M. y Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nature methods*, 10(10), pp.957–963.
- Marraffini, L.A. y Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 322(5909), pp.1843–1845.
- Marraffini, L.A. y Sontheimer, E.J., 2009. Invasive DNA, Chopped and in the CRISPR. *Structure*, 17(6), pp.786–788.
- Mendeley Desktop, programa de ordenador.
<http://www.mendeley.com/>
- Mojica, F.J., Juez, G. y Rodríguez-Valera, F., 1993. Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular microbiology*, 9(3), pp.613–621.
- Mojica, F.J.M. et al., 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, 17(1), pp.85–93.
- Mojica, F.J. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, 36(1), pp.244–246.
- Mojica, F.J.M. et al., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, 60(2), pp.174–182.
- Mojica, F.J.M. et al., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3), pp.733–740.
- Mojica, F.J.M. y Garret, R.A., 2013. Discovery and Seminal Developments in the CRISPR Field. En *CRISPR-Cas Systems RNA-Mediated Adaptive Immunity in Bacteria and Archaea*. Barrangou, R. y van der Oost, J., eds. Springer, p. 299.
- Nakata, A., Amemura, M. y Makino, K., 1989. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *Journal of Bacteriology*, 171(6), pp.3553–3556.
- Nakayama, T. et al., 2013. Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus tropicalis*. *Genesis*, 51(12), pp.835–843.
- Nishimasu, H. et al., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), pp.935–949.
- Paix, a. et al., 2014. Scalable and Versatile Genome Editing Using Linear DNAs with Microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics*, 198(December), pp.1347–1356.

Pascual Calaforra, L.F. y Latorre Castillo, A., 2013. Ingeniería genética. En *Genética conceptos esenciales*. Benito, C. y Espino, F.J. eds. Panamericana, p. 588.

Perez-Pinera, P. et al., 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10(10), pp.973–6.

Pourcel, C., Salvignol, G. & Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3), pp.653–663.

PubMed. National Center for Biotechnology Information, U.S. National Library of Medicine (Bethesda, MD, US).
<http://www.ncbi.nlm.nih.gov/pubmed>.

Ran, F. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–308.

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6).

Rees, D.C., Williams, T.N. & Gladwin, M.T., 2010. Sickle-cell disease. *The Lancet*, 376(9757), pp.2018–2031.

Sakuma, T. y Yamamoto, T., 2015. CRISPR/Cas9: The Leading Edge of Genome Editing Technology. In T. Yamamoto, ed. *Targeted Genome Editing Using Site-Specific Nucleases*. Tokyo: Springer, p. 205.

Sakuma, T. et al., 2014. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific reports*, 4, p.5400.

Sander, J.D. y Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*, 32(4), pp.347–55.

Sapranauskas, R. et al., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research*, 39(21), pp.9275–9282.

Sasaki, H. et al., 2014. CRISPR/Cas9-mediated gene knockout in the ascidian *Ciona intestinalis*. *Development, Growth & Differentiation*, 56(7), pp.499–510.

Scharenberg, A.M., Duchateau, P. & Smith, J., 2013. Genome engineering with TAL-effector nucleases and alternative modular nuclease technologies. *Current Gene Therapy*, 13, pp.291–303.

Schwank, G. et al., 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*, 13(6), pp.653–658.

Shao, Y. et al., 2014. CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. *Nature protocols*, 9(10), pp.2493–2512.

Shen, B. et al., 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods*, 11(4), pp.399–402.

- Steinberg, M.H. y Sebastiani, P., 2012. Genetic modifiers of sickle cell disease. *American Journal of Hematology*, 87(8), pp.795–803.
- Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), pp.62–7.
- Stoddard, B.L., 2011. Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure*, 19, pp.7–15.
- Tang, T.H. et al., 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(11), pp.7536–7541.
- Tsai, S.Q. et al., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*, 32(6), pp.569–576.
- United Nations, 1992. Convention on biological diversity. p. 30..
- Urnov, F.D. et al., 2010. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11, pp.636–646.
- Wang, H. et al., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, 153, pp.910–918.
- Wang, Y. et al., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, 32(9), pp.947–951.
- Xie, K. y Yang, Y., 2013. RNA-Guided genome editing in plants using a CRISPR-Cas system. *Molecular Plant*, 6(6), pp.1975–1983.
- Yamamoto, T., 2015. *Targeted Genome Editing Using Site-Specific Nucleases ZFNs, TALENs, and the CRISPR/Cas9 System*. Springer, pp. 205.
- Yang, H., Wang, H. y Jaenisch, R., 2014. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nature Protocols*, 9(8), pp.1956–1968.
- Yin, H. et al., 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology*, 32(6), pp.551–3.
- Yosef, I., Goren, M.G. y Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research*, 40(12), pp.5569–5576.
- Zhang, F., 2014. CRISPR-Cas systems and methods for altering expression of gene products. *US Patent 8,697,359*, p.150.